Fall 2019

Spatiotemporal Downscaling Rainfall Predictions of North American Regional Climate Change Assessment Program for Entire Virginia

Zhaoyi Cai
Old Dominion University, zcai001@odu.edu

SPATIOTEMPORAL DOWNSCALING RAINFALL PREDICTIONS OF NORTH AMERICAN REGIONAL CLIMATE CHANGE ASSESSMENT PROGRAM FOR ENTIRE VIRGINIA

Zhaoyi Cai
B.E. August 2017, Southeast University, China

A Thesis Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

CIVIL ENGINEERING

OLD DOMINION UNIVERSITY

November 2019

Approved by:

Dr. Xixi Wang (Director)

Dr. Isao Ishibashi (Member)

Dr. Mujde Erten-Unal (Member)

ABSTRACT

SPATIOTEMPORAL DOWNSCALING RAINFALL PREDICTIONS OF NORTH AMERICAN REGIONAL CLIMATE CHANGE ASSESSMENT PROGRAM FOR ENTIRE VIRGINIA

Zhaoyi Cai
Old Dominion University, 2019
Director: Xixi Wang

This thesis developed a statistical downscaling approach, which consists of a series of linear regression equations, to spatiotemporally downscale the rainfall predictions from North American Regional Climate Change Assessment Program (NARCCAP) in accordance with the 15-min observed rainfall data at the rain gauges across the state of Virginia. NARCCAP has generated twelve different region-global climate models (RCM-GCMs) with a temporal resolution of 3 hr and a spatial resolution of 50 km over the entire America. Although it has been downscaled already, such resolutions are still too coarse to represent the rain gauges. This means that the RCM-GCMs' predictions need to be further downscaled for watershed planning and management as well as hydrologic engineering design. In this regard, this thesis developed the statistical downscaling approach using the predictions by one of the RCM-GCMs and then validated the applicability of the approach using the predictions by the other RCM-GCMs. The development and validation were implemented by comparing the RCM-GCMs' predictions with the observed data. Future studies should better utilize the predictions of all twelve RCM-GCMs and try some nonlinear algorithms to minimize either underestimating or overpredicting some extreme rainfalls for a duration of longer than 3 hr.

This thesis is dedicated to the proposition that the harder you work, the luckier you get.

ACKNOWLEDGMENTS

I would like to thank my academic advisor, Dr. Xixi Wang, for his guiding me through the development of this thesis and problem-solving using Storm Water Management Model (SWMM). Dr. Wang has been very patient throughout the semesters during which I have been doing my best to complete this thesis. I could not have accomplished this work without his help and support. In addition, I want to extend my appreciation to my committee members, Dr. Isao Ishibashi and Dr. Mujde Erten-Unal, for their careful review and evaluation of this thesis. Further, I want to express my heartful gratitude to my parents, Zhiqiong Zhao (mother) and Hang Cai (father), for their selfless and persistent support throughout my life.

TABLES OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

PREFACE

In the past century, the concentration of greenhouse gases (GHGs), including CO_2, in the Earth's atmosphere has been rising due to increasing industrial activities. This has caused, and will continue to cause, large-scale variations in atmospheric processes, which in turn may lead to the decrease in precipitation amount but increase in extreme rainfall intensity, and atmosphere temperature. The overarching goal of this thesis was to advance our understanding of non-stationarity resulting from climate change. The specific objective was to downscale the predictions of various region-global models (RCM-GCMs) ensembles for the state of Virginia, generating extreme rainfall datasets for developing probability-based rainfall intensity-duration-frequency (IDF) curves, which allow us to take into account influences of the non-stationarity in designing hydraulic structures, increasing the structures' resilience while avoiding either over or under committing resources.

This thesis is organized into five chapters. *Chapter 1 (Introduction)* describes the problem and challenge from climate change and puts forward the thesis goal and study objective. *Chapter 2 (Literature Review)* overviews the historical development, evolution, and application of various RCM-GCMs, summarizes the commonly used downscaling methods, and highlights the research needs. *Chapter 3 (Methods)* describes the methods used by this thesis study. *Chapter 4 (Results and Discussion)* presents the results of this thesis and discusses their validity and applicability. *Chapter 5 (Conclusions and Recommendations)* concisely summarizes the major findings from this thesis study and makes recommendations for future studies.

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

An intensity-duration-frequency (IDF) curve is a mathematical function that relates rainfall intensity with its duration and frequency of occurrence. Quantifying extreme rainfall characteristics (i.e. intensity, duration, and frequency) is imperative in hydrologic engineering design. In this regard, for a given duration and return period, the rainfall intensity can be determined from the rainfall IDF curve. In practice, IDF curves have also been widely used as an effective tool in approximating extreme rainfall for storm water management and other water-related infrastructure designs.

In the past century, the concentration of greenhouse gases (GHGs), including CO_2, in the Earth's atmosphere has been rising due to increasing industrial activities. This has caused, and will continue to cause, large-scale variations in atmospheric processes, which in turn may lead to the decrease in precipitation amount but increase in extreme rainfall intensity, and atmosphere temperature. Such changes in rainfall characteristics can invalidate the existing IDF curves, which were developed by assuming non-stationary climates, leading to either over or under committing resources and an unknown risk in designing infrastructures. Current design standards are mainly based on the IDF curves developed using historic climate information. For example, a dam that is designed to control a 100-year flood may provide a significantly lower level of protection if the intensity and duration of the 100-year storm event increases. To adapt to future climates, it is imperative to update the current standards for water infrastructure design and to evaluate possible influences of climate change on the existing IDF curves.

Changes in the hydrologic cycle resulting from greenhouse gases have been projected to cause

variations in rainfall intensity, duration, and frequency. Incorporating potential effects of climate change into engineering design is practically needed to reduce uncertainty and increase resilience of infrastructure. In this regard, revisiting and updating the existing IDF curves for future climate scenarios are raising more attention.

In terms of how extreme rainfall characteristics would be modified in the future, Golbahar *et al.* (2013) found that a large uncertainty of the projected rainfall intensities by six climate models for long (> 4 h) durations makes it difficult to draw any strong conclusions about the expected changes on future rainfall intensity in Alabama. The inconsistent results can be attributed to the difference in 1) physical parameterizations, especially of radiative and precipitation-forming processes, among different General Circulation Models (GCMs) and Regional Climate Models (RCMs); and 2) initial and boundary conditions for each climate projection. Given the large uncertainty in the output from paired RCMs-GCMs, performing an uncertainty analysis and creating probability based IDF curves has been considered as an emerging research need. Alain *et al.* (2007) compared the CRCM-simulated annual May to October maximum rainfall depth (MOAM) series (2-, 6-, 12-, and 24-h durations) to the available historical records over Southern Quebec, Canada. Those authors found that increases in regional MOAM values can be detected between the control and future periods and that uncertainties in these changes tend to increase with increase of return period and/or duration. Thus, multi-model ensembles will need to be analyzed to quantify such uncertainties. Similarly, Emori *et al.* (2005) showed that the simulation of extreme daily precipitations was highly dependent on model parameterization, implying that a model with a higher resolution does not necessarily result in better predictions of future climate than a model with a lower resolution. These ensembles must include the use of outputs of various combinations of the GCMs and RCMs to take into account possible future scenarios, which

represent how socioeconomic developments, technology advancements, and greenhouse gas emissions would be conceptualized and mathematically described.

1.2 Objective

The overarching goal of this thesis was to advance our understanding of non-stationarity resulting from climate change. The specific objective was to downscale the predictions of various RCM-GCMs ensembles for the state of Virginia, generating extreme rainfall datasets for developing probability-based IDF curves, which allow us to take into account influences of the non-stationarity in designing hydraulic structures, increasing the structures' resilience while avoiding either over or under committing resources.

CHAPTER 2

LITERATURE REVIEW

Predictions of future climates have been extensively studied and well documented in existing literature. Some of the existing documents focus on the development and validation of various prediction models, while others focus on applications of the predicted results. Given the large amount of such documents, this chapter was carved by following the overview presented by Africa and Latin American Resilience to Climate Change Project (2014) from the United States Agency for International Development, with significant enrichments from other sources and the author's insights. Its main purpose was to identify what has been done and what needs to be done, highlighting the research necessity and contributions of this thesis.

2.1 General Circulation Model or Global Circulation Models (GCMs)

Decision makers are increasingly demanding climate information at the national to local scale in order to address the risk posed by projected climate changes and their anticipated impacts. To respond to the needs of decision makers to plan for climate change, a variety of reports, tools, and datasets provide projected climate impacts at spatial and temporal scales much finer than those at which the projections are made. It is important to recognize the variety of assumptions behind the techniques used to derive such information and the limitations they impose on the results. The main tools used to project climate are General Circulation Models (GCMs), which are computer models that mathematically represent various physical processes of the global climate system. General or global circulation models (GCMs) simulate the Earth's climate via mathematical equations that describe atmospheric, oceanic, and biotic processes, interactions, and feedbacks. They are the primary tools that provide reasonably accurate global-, hemispheric-

, and continental-scale climate information and are used to understand present climate and future climate scenarios under increased greenhouse gas concentrations.

A GCM is composed of many grid cells that represent horizontal and vertical areas on the Earth's surface. In each of the cells, GCMs compute the following: water vapor and cloud atmospheric interactions, direct and indirect effects of aerosols on radiation and precipitation, changes in snow cover and sea ice, the storage of heat in soils and oceans, surfaces fluxes of heat and moisture, and large-scale transport of heat and water by the atmosphere and oceans.

The spatial resolution of GCMs is generally quite coarse, with a grid size of about 100–500 kilometers. Each modeled grid cell is homogenous, (i.e., within the cell there is one value for a given variable). Moreover, they are usually dependable at temporal scales of monthly means and longer. In summary, GCMs provide quantitative estimates of future climate change that are valid at the global and continental scale and over long periods.

Although GCMs are valuable predictive tools, they cannot account for fine-scale heterogeneity of climate variability and change due to their coarse resolution. Numerous landscape features such as mountains, water bodies, infrastructure, land-cover characteristics, and components of the climate system such as convective clouds and coastal breezes, have scales that are much finer than 100–500 kilometers. Such heterogeneities are important for decision makers who require information on potential impacts on crop production, hydrology, species distribution, etc. at scales of 10–50 kilometers.

2.2 Downscaling Methods

Readily available climate change projections are provided at global and continental spatial scales for the end of the 21st century (Intergovernmental Panel on Climate Change [IPCC], 2007). These projections, however, do not fit the needs of sub-national adaptation planning that requires

regional and/or local projections of likely conditions five to 10 years from now. Moreover, decision makers are interested in understanding the impacts of climate change on specific sectors, e.g., agricultural production, food security, disease prevalence, and population vulnerability.

In response to this demand, numerous impact and vulnerability assessments produced at different scales, from global to local, provide climate change impact results at spatial scales much finer than those at which projections are initially made. To produce such results, combinations of methods and indicators are often used, each with its own assumptions, advantages, and disadvantages. In reports, these essential factors may not be adequately communicated to the reader, thus leaving him/her without the ability to understand potential discrepancies between different reports. Often, global or continental-scale information is directly used to produce local-scale impact maps, which is not appropriate since this large-scale information does not account for differences at the local scale.

Downscaling is a technique that is used to extract high-resolution information from regional scale variables produced by coarse resolution models. Any information that is presented at spatial scales finer than 100 kilometers x 100 kilometers and temporal scales finer than monthly values has undergone a process called downscaling. While it produces climatic information at scales finer than the initial projections, this process involves additional information, data, and assumptions, leading to further uncertainties and limitations of the results, a consequence that is often not made explicit to end-users. International organizations or national governments currently provide no official guidance that assists researchers, practitioners, and decision makers in determining climate projection parameters, downscaling methods, and data sources that best meet their needs. Since the research community is still developing downscaling methods, users

often need to read highly technical and specialized explanations in order to understand and adequately apply the results for impact studies, planning, or decision-making.

The followings are important considerations and recommendations to keep in mind when designing and interpreting fine-scale information on climate change and its impacts:

1) Downscaling relies on the assumption that local climate is a combination of large-scale climatic/atmospheric features (global, hemispheric, continental, regional) and local conditions (topography, water bodies, land surface properties). Representation of the latter is generally beyond the capacity of current GCMs.

2) Deriving climate projections at local scales is a multistep process and at each step, assumptions and approximations are made. Uncertainties are inherent in projections of changes in climate and their impacts. They arise from different sources and need to be kept in mind, whether explicitly quantified or not.

3) Downscaling can be applied spatially and temporally. Oftentimes, several downscaling methods are combined to obtain climate change information at desired spatial and temporal scales.

There are two principal ways to combine the information on local conditions with large-scale climate projections:

Dynamical: by explicitly including additional data and physical processes in models similar to GCMs but at a much higher resolution and covering only select portions of the globe1. This method has numerous advantages but is computationally intensive and requires large volumes of data as well as a high level of expertise to implement and interpret results, often beyond the capacities of institutions in developing countries.

Statistical: by establishing statistical relationships between large-scale climate features that GCMs and local climate characteristics provide. In contrast to the dynamical method, the statistical methods are easy to implement and interpret. They require minimal computing resources but rely heavily on historical climate observations and the assumption that currently observed relationships will carry into the future. However, high quality historical records often are not available in developing countries.

2.3 Dynamical Downscaling

Global climate models (GCMs) with the ability to capture large-scale circulations are useful tools for climate simulation. RCMs take the large-scale atmospheric information supplied by GCM output at the lateral boundaries and incorporate more complex topography, the land-sea contrast, surface heterogeneities, and detailed descriptions of physical processes in order to generate realistic climate information at a spatial resolution of approximately 20 to 50 kilometers. However, GCMs have limited suitability in representing regional climate variability, especially for regions with complex terrain, owing to the coarse resolution and simple physical parameterizations used in GCMs. The dynamical downscaling technique consists mainly of nesting high-resolution regional climate models (RCMs) into GCMs or using reanalysis data within a limited area of interest. RCMs apply higher resolution topography, the land-sea contrast, surface heterogeneities, and finer physical processes to simulate climate more accurately.

Since the RCM is nested in a GCM, the overall quality of dynamically downscaled RCM output is tied to the accuracy of the large-scale forcing of the GCM and its biases. Despite recovering important regional-scale features that are underestimated in coarse-resolution GCMs, RCM outputs are still subject to systematic errors and therefore often require a bias correction as well as further downscaling to a higher resolution.

2.3.1 Regional Climate Models (RCMs)

An RCM is like a GCM but has higher resolution and additional regional information, which enables it to better represent local landscape and possibly local atmospheric processes. The global model simulates the response of the global circulation to changes in atmospheric composition through a large number of processes, but some of them need to be approximated due to the coarse resolution of the models. On the other hand, at the resolution of 25–50 km for portions of the globe, the RCM is able to capture some of those smaller-scale processes more realistically. Atmospheric fields (e.g., surface pressure, wind, temperature, and humidity) simulated by a GCM are fed into the vertical and horizontal boundaries of the RCM. Locally specific data and physics-based equations are then used to process this information and obtain regional climate outputs. The primary advantage of RCMs is their ability to model atmospheric processes and land cover changes explicitly. Although there has been great advancement during the past decade in the technical ability of RCMs to simulate regional climate, significant challenges and concerns still exist. Since smaller grid cells, more surface information, and often more processes are included in an RCM, the number of computations might be as large, if not larger, than in a GCM that covers the entire globe4. Thus, RCMs are computationally demanding and may require as much processing time as a GCM to compute projections (Wilby et al., 2009). They also require a substantial amount of input, e.g., surface properties and high-frequency GCM information. In addition, complex calibration procedures are often needed to make realistic simulations.

Just like GCMs, RCMs have difficulty accurately simulating convective precipitation, which is a major concern for tropical regions. Most RCMs also do not accurately simulate extreme precipitation — a systematic bias that can worsen as the resolution is increased. Statistical bias

corrections often need to be performed to better match the model output to the observations (Brown et al., 2008). In some cases, fine adjustments to the convective schemes can improve the realism of simulated rainfall, but these adjustments require substantial expertise and reduce geographic portability — that is, they create a version of the model that is well adjusted to a particular region but that may perform poorly elsewhere.

The quality of RCM results also depends on the driving GCM information. For example, if the GCM misplaces storm tracks, there will be errors in the RCM's precipitation climatology (Wilby et al., 2009). Additionally, different RCMs contain distinct dynamical schemes and physical parameters, which means that RCMs driven by the same GCM can produce different results.

Finally, the grid-box size of an RCM is typically greater than 10 kilometers, which is still too coarse for hydrological and agricultural impact studies that require more local- or station-scale climate information (Benestad, 2009). To obtain higher resolution results, statistical methods are used in lieu of RCMs, or RCM output is further downscaled via statistical means.

Generally, the RCMs are being adopted worldwide as they offer many advantages such as (i) giving intense events that will be smoothened in coarse resolution but may still miss the most extremes; (ii) giving phenomenological values diurnal cycle; (iii) having more numerical stability and accuracy as these cover only a fraction of the globe and require short-time steps; (iv) providing improvement in climate simulations, especially for precipitation that has high spatial variability; and (v) boundary conditions based on actual observations that provide information on fine-scale climate behavior besides isolating GCM error from the errors intrinsic to RCM. The RCMs have some limitations too: for example, (i) simulation is dependent upon the boundary conditions supplied from other source; (ii) climate needs parameterization for sub grid-scale processes, surface atmosphere coupling, and radiation transfer and cloud microphysics; (iii) only

a limited number of scenario runs are available, and the "time slice" approach is used; (iv) the outputs of the dynamically downscaled RCM are dependent on the precision of the forcing GCM and its unfairness (Seaby et al., 2013); (v) the outputs of the RCM are liable to systematic errors and may involve a method of bias removal in addition to downscaling techniques for higher resolution; and (vi) climate information in grid-box size of an RCM is at higher resolution (larger than10 km) than local- or station-scale, which make such downscaling inappropriate for hydrologic and agricultural impact studies (Benestad, 2009).

RCMs are developed by research institutions that have sufficient computational capacity and technical expertise. Various RCMs differ in their numerical, physical, and technical aspects. The most commonly used RCMs in climate change downscaling studies include the U.S. Regional Climate Model Version 3 (RegCM3); Canadian Regional Climate Model (CRCM); UK Met Office Hadley Centre's Regional Climate Model Version 3 (HadRM3); German Regional Climate Model (REMO); Dutch Regional Atmospheric Climate Model (RACMO); and German HIRHAM, which combines the dynamics of the High Resolution Limited Area Model (HIRLAM) and European Centre-Hamburg (ECHAM) models.

Although the above models have been developed primarily over North America and Europe, they can be adapted to any region of the globe by incorporating appropriate information on terrain, land-cover, hydrology, and so on; hence, several RCMs can be used over a given region. However, downscaled results can differ depending on which RCM(s) is used. It is important to recognize that a single RCM will most likely not provide 'accurate' results.

In this thesis, we chose six different RCMs from the North American Regional Climate Change Assessment Program (NARCCAP). The general NARCCAP strategy (as in most RCM applications) consists of two phases. In Phase I, six RCMs were forced with global reanalysis

from the National Center for Environmental Prediction/Department of Energy (NCEP/DOE) reanalysis as the boundary conditions. Because the reanalysis effectively consists of weather prediction model analysis fields (with "frozen" model versions and analysis systems), it is appropriate to compare the RCM output with observations on a time step basis. In the second phase, GCM output was used to provide boundary conditions for both historic and future climate runs. For the historic run, given the chaotic nature of the atmosphere as represented in the GCM boundary conditions, comparisons with observations is only possible in a statistical context. The six RCMs participating in NARCCAP are the Hadley Regional Model 3 (HRM3), the Regional Climate Model version 3 (RCM3), the Canadian Regional Climate Model (CRCM), the NCEP Experimental Climate Prediction Center Regional Spectral Model (ECPC), the MM5-PSU/NCAR mesoscale model (MM5), and the Weather Research and Forecasting model (WRF). In Phase I of NARCCAP, 25-year (1980-2004) RCM simulations were implemented using the NCEP/DOE Reanalysis for boundary conditions. In Phase II, each RCM was nested within at least one GCM at 50 km spatial resolution for the periods 1971-2000 and 2041-2070. In this thesis, all 12 combinations of RCMs and GCMs for Phase II had been archived and all of the RCM-based analyses used seasonal and annual mean precipitation (P) derived from 3-hour NARCCAP output.

2.3.2 Approaches Developments

Coarse-grid GCM simulation output is used for initial and lateral boundary conditions, known as "one-way nesting approach" (Mearns et al., 2003). The "nested" RCM approach was first applied in climate change studies in the late 1980s by Dickinson, Errico, Giorgi, and Bates (1989). However, most researches focused more on improving the technical ability of RCMs to simulate regional climate in the next 20 years. Even though one-way nested approaches have generally

been used in many RCM simulations, feedback from RCM to GCM could still hardly be found until the 21st century. Two-way nesting means an interactive numerical model integration, where a part of the integration area is computed at a finer horizontal resolution than the coarser resolved residual area. The basic idea behind this technique is the reduced computing time compared to an integration at the finer resolution over the whole integration area. Two-way nesting techniques have already been applied between atmospheric limited area models at different horizontal resolutions (Phillips and Shukla, 1973). Lorenz and Jacob (2005) developed a two-way nested ECHAM4-REMO atmospheric climate model system, integrated numerically stable for a 10-year period using a two-way nesting region and found a positive influence on the simulation of the global climate, even in regions not covered by the two-way nesting domain. Also, they found that the systematic error can be reduced globally by a more detailed representation of this particular region. Bowden et al. (2012) compared the three nudging techniques in the WRF model using two-way nesting to determine the influence of interior nudging on mean error. Jeon et al. (2019) also implemented an online two-way nesting framework to improve global surface tides in the Hybrid Coordinate Ocean Model (HYCOM). In all this research, high-resolution child domains are coupled with relatively low-resolution parent domains for computational efficiency.

2.4 Statistical Downscaling

Statistical downscaling involves the establishment of empirical relationships between historical and/or current large-scale atmospheric and local climate variables. Once a relationship has been determined and validated, future atmospheric variables that GCMs project are used to predict future local climate variables. Statistical downscaling can produce site-specific climate projections, which RCMs cannot provide since they are computationally limited to a 20 to 50

kilometers spatial resolution. However, this approach relies on the critical assumption that the relationship between present large-scale circulation and local climate remains valid under different forcing conditions of possible future climates. It is unknown whether present-day statistical relationships between large- and regional-scale variables will be upheld in the future climate system. The main advantage of statistical downscaling methods is that they are computationally inexpensive and appropriate when computational resources are limited. Regression-based downscaling is a widely applied method in practice. It formalizes mathematically the relationship between large-scale predictors and the small-scale predictand. Because of its much lower computational cost, statistical downscaling is almost always used for multi-model downscaling. Statistical downscaling relies on empirical mathematical relationships to go from large-scale predictors to fine scale predictands. These relationships are often much faster to apply than dynamical downscaling, which makes them ideal for downscaling large ensembles of GCMs for multiple time periods or scenarios. However, they are subject to the stationarity assumption that the relationship between the predictors and predictands continues to hold, even in a changed climate. Although statistical models are valuable tools for downscaling multi-model ensembles, they do not produce a full complement of variables like dynamical downscaling.

Oftentimes, dynamical and statistical approaches are used in conjunction. Dynamical-statistical downscaling involves the use of an RCM to downscale GCM output before statistical equations are used to further downscale RCM output to a finer resolution. Dynamical downscaling improves specific aspects of regional climate modeling and provides better predictors for further statistical downscaling to higher-resolution output. Statistical-dynamical downscaling is a somewhat more complex approach but is less computationally demanding in comparison to

dynamical downscaling. This method statistically pre-filters GCM outputs into a few characteristic states that are further used in RCM simulations.

Statistical downscaling consists of a heterogeneous group of methods that vary in sophistication and applicability. They are all relatively simple to implement but require a sufficient amount of high-quality observational data. Most statistical downscaling methods can be classified into three main categories: Perfect Prognosis, Model Output Statistics and Weather Generators.

2.4.1 Approaches Developments

The basic premise behind Perfect Prognosis (PP) approaches to statistical downscaling is that GCMs are able to simulate the large-scale atmospheric climate fields realistically, even if fields with high spatial variability, like precipitation, are poorly simulated. Thus, statistical relationships are sought with variables in which there is high confidence, while ignoring those in which there is low confidence. Most PP approaches disregard any residual noise term although some newer PP approaches explicitly provide a noise model to help capture the variability and extremes. Approaches that include a noise model are often referred to as stochastic, while those that do not are termed deterministic.

Building a PP downscaling scheme requires two steps that are often performed together. They are the identification of suitable, observed large-scale predictors, and the development of the statistical relationship between them and the local-scale observations. It is important that predictors that capture the effects of climate change are included in the scheme if it is to be used to downscale future projections. This consideration needs to be kept in mind when identifying predictors based on historical time series that may contain only a small climate change signal. In general, the predictor choice will vary depending on the region and season. Various large-scale predictors for downscaling precipitation have been explored in Wilby and Wigley (2000), along

with a comparison of the observed and simulated fields. Often these predictors are high-dimensional fields of grid-based values. Since these fields frequently have high levels of spatial correlation, the grid-point values are not independent. Thus, it is relatively common to reduce the dimensionality of the predictor field in some way. Common techniques for this include principal component analysis (Hannachi, 2007), canonical correlation analysis (Hertig and Jacobeit, 2008; Palatella et al., 2010), maximum covariance analysis (Tippett et al., 2008), support vector machine (Nayak and Ghosh, 2013), and physically-motivated transformations such as using an ENSO index or weather types (Wu et al., 2010). Weather types are circulation patterns or regimes that occur frequently in a location. They can be defined subjectively, by visually inspecting synoptic maps, or objectively using clustering and classification algorithms.

There are many ways to establish the statistical relationship between the predictors and predictands in PP, though in each case the relationship is calibrated using observed variables before being applied to climate model output. Each statistical model has its own set of assumptions and level of complexity. Among the simplest models are linear regression models. These assume that the variables involved are Gaussian-distributed, which is not true of precipitation fields on short timescales, including daily. This assumption has been relaxed in the framework of the generalized linear model. Also, the linearity dependence has been replaced by non-parametric smooth functions in the generalized additive model (e.g., Vrac et al., 2007). All of these methods focus on predicting the mean conditional on a set of predictors. In order to quantify the variance (or higher-order moments) dependence on a set of predictors, vector generalized linear models can be used.

Several non-linear regression techniques have also been applied to the statistical downscaling problem. Such techniques include the application of artificial neural networks to downscale

precipitation (e.g., Haylock et al., 2006). Another method that has been applied to downscaling is the analogue method (e.g., Zorita and von Storch, 1999). In this method, a selected metric is used to identify the most similar situation in the historical record and the corresponding local observations are used as the prediction. One major limitation of this approach is that it cannot produce precipitation amounts that have not been observed in the past.

The many different statistical models that can be used in PP downscaling make various assumptions and have various limitations. All PP approaches do, however, share two major assumptions. Firstly, that suitable predictors are well-simulated by the GCM; that is, only fields that have been evaluated and found to perform well should be used. Secondly, that the relationship identified between the predictor and predictands is stationary. That is, it is assumed that, although the climate changes, the relationship between the identified variables does not change.

2.4.2 Model Output Statistics

Unlike PP techniques, Model Output Statistics (MOS) methods develop statistical relationships between simulated predictors and observed predictands. They are most often applied to climate model-simulated fields of the same variable being predicted. That is, a MOS method can be used to correct the RCM-simulated precipitation field, in order to account for the difference between areal-gridded means and local point observations of precipitation. As such, MOS methods can often be thought of as statistical corrections to RCM-simulated outputs; indeed, they have been used in numerical weather prediction (NWP) for some time (Glahn and Lowry, 1972; Kalnay, 2003).

If the RCM simulation is driven by an atmospheric re-analysis, then there is a direct correspondence between the simulation and observations. In this case, the MOS method will

relate the simulated and observed time series through regression techniques. If, on the other hand, the RCM simulation is not driven by a re-analysis, then this direct relationship does not exist between the simulation and observations. In this case, the MOS methods can only be used to link the distributions of the variables.

At their simplest, MOS methods provide a bias correction of the present-day simulated field to match the observations. For variables such as temperature, this is usually a simple arithmetic (e.g., additive) correction while, for precipitation, this is applied as a scaling factor, often calculated and applied separately for each month or season. A more complex approach is quantile matching. In this approach, different intensities are considered individually such that the simulated cumulative distribution function is adjusted to match the observed cumulative distribution function. Similar bias-correction approaches have been further developed to account for persistence in the precipitation fields (Johnson and Sharma, 2012).

2.4.3 Weather Generators

Weather generators are statistical models that produce random sequences of climate variables with statistical properties that match those of the observed variables. Weather generators were not originally developed with spatial downscaling but are typically used in temporal downscaling. For example, they are used to generate daily sequences of weather variables (e.g., precipitation, maximum and minimum temperature, humidity, etc.) that correspond to monthly or annual averages or amounts. Temporal downscaling is necessary for some impact models that require local spatial data at a daily resolution, which GCMs cannot reliably provide. Weather generators produce sequences of daily values, but since different weather sequences may be associated with a given set of, for example, monthly values, multiple sequences commonly are generated to be further used in impact models. This method usually generates very long time series to assist in

the study of floods, planning for large water engineering projects, and so on. As downscaling tools, they are often used with the statistical relationships being developed from observed data. The same statistical properties are calculated on both a present-day and future climate simulation. The simulated changes in these properties are then applied to the observed parameters (Semenov and Barrow, 1997).

2.5 Summary

In most cases, a sequence of different methods is needed to obtain results at the desired resolution; however, the analysis of select reports presenting changes in climate and/or their impacts has shown the following points:

1) Information on downscaling and the limitations of the results are often not appropriately highlighted, leading the user to believe that the results are "true" and valid at the resolution presented. Extensive reading of technical documentation is often needed to uncover all the steps and assumptions that led to the final results.

2) Uncertainties inherent in projections and additionally arising from applied downscaling are often not presented, quantified, nor discussed, leading the user to interpret the numerical results at face value.

3) Validation of downscaled results (on historical data) is often omitted; comparing downscaled results to high-resolution observed information would highlight systematic biases and the limitations of results.

The above deficiencies most frequently result from simple oversight by the authors of the report or their efforts to make it easy to use. However, they are important, and an expert user may be able to detect them and estimate the limitations of the results. The overall diversity of the

approaches and methods in existing reports and publications reflects the diversity of the goals and resources of each assessment. Thus, there is no single best downscaling approach, and downscaling methods will depend on the desired spatial and temporal resolution of outputs and the climate characteristics of the highest impact of interest. In light of current approaches and practices reviewed in this report, it is possible to make the recommendations that follow:

1) When designing assessments of climate change and its impacts at sub-regional scales, a thorough evaluation of the information needs and the relevance of existing information should be carried out first. If the need for an original downscaling of the projections is confirmed, the approach should be selected based on the information needs and also, importantly, on available resources (data, computing resources, expertise, and timeframes).

2) When using/interpreting existing results/reports, the coarse resolution of the initial projections and the scales at which they are valid need to be kept in mind. Any results presenting fine-scale spatial details or using high temporal resolution data have undergone a manipulation (usually a sequence of manipulations) of the original projections, whether this process is described or not. It is only through an evaluation of the employed downscaling procedure that the validity of the results at a fine resolution and the value added over initial coarse projections can be assessed. Results that look detailed may actually not be robust; in general, a rigorous downscaling process requires including additional information, and a simple interpolation from coarse- to fine-scale may not lead to reliable results. Therefore, it is important to understand (and research if not directly available) at least the broad aspects of the applied downscaling.

3) Since uncertainty is inherent to the projections, an estimate of it — quantitative or at least qualitative — should always be included and carried through the downscaling process. Such

an estimate should at least include different potential future climate states and ideally should also estimate the influence of the downscaling procedure on the results.

CHAPTER 3

METHODS

3.1 Study Site

This thesis was conducted for the entire state of Virginia, most of which is located within the Chesapeake Bay Watershed. The state has a humid subtropical climate, with an annual average of 35 to 45 days of thunderstorm activity and an average annual precipitation of 1080 mm. The average temperature varies from -3°C in January to 30°C in July.

3.1.1 Site Characteristics

Virginia has a significant topographic relief, with elevations varying from Virginia Beach in the east at sea level to Mount Rogers in the west at 1746 m above sea level. The major gradations occur at the edges of the Atlantic Ocean, the end of the Piedmont, and the Blue Ridge and Allegheny chains of the Appalachian Mountains. The moderating influence of the ocean from the east, powered by the Gulf Stream, also creates the potential for hurricanes near the mouth of Chesapeake Bay. Cold air masses arrive over the mountains, especially in winter, which can lead to significant snowfalls when coastal storms, known as nor'easters, move up the Atlantic coast. The interaction of these elements with the state's topography creates micro-climates in the Shenandoah Valley, the mountainous southwest, and the coastal plains that are slightly but noticeably distinct from each other. To differentiate and characterize the micro-climates, NOAA (National Oceanic and Atmospheric Administration) subdivides Virginia into six climatic divisions, namely Tidewater, Eastern Piedmont, Western Piedmont, Northern, Central Mountain, and Southwestern Mountain. On the other hand, to differentiate and characterize the topographic relief, USGS (U.S. Geological Survey) subdivides Virginia into five physiographic divisions, namely Coastal Plain, Piedmont, Blue Ridge, Valley and Ridge, and Appalachian Plateaus. The

Coastal Plain division is between the Atlantic coast and the fall line. It includes the Eastern Shore and major estuaries of Chesapeake Bay. The Piedmont division is a series of sedimentary and igneous rock-based foothills east of the mountains which were formed in the Mesozoic era. The region, known for its heavy clay soil, includes the Southwest Mountains around City of Charlottesville. The Blue Ridge division is a physiographic province of the Appalachian Mountains with the highest points in the state, the tallest being Mount Rogers. The Valley and Ridge division is west of the mountains and includes the Great Appalachian Valley. The region is carbonate rock based and includes Massanutten Mountain. The Cumberland Plateau and the Cumberland Mountains are in the southwest corner of Virginia, south of the Allegheny Plateau (Figure 3.1).

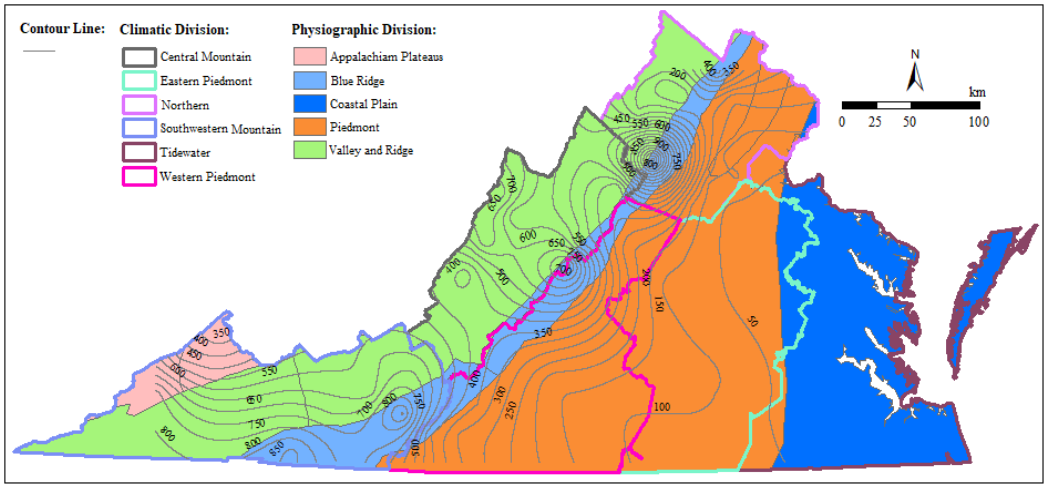

Figure 3.1. The Virginian climatic and physiographic divisions superimposed by the topographic elevation contours at a 50-m interval.

3.1.2 Rain Gauges

This study overlaid the climatic and physiographic divisions, resulting in 13 zones (Figure 3.2 and Table 3.1), within which the 57 rain gauges with 15-min rainfall data are located. Note that there is no such rain gauge in one of the zones (i.e., CPZ03). This project did some analyses by pooling together the rainfall data at the rain gauges within each of the climatic-physiographic zones, as detailed in the following contexts.

Table 3.1. The climatic-physiographic zones with inclusive rain gauges.

Climatic-Physiographic Zone	Climatic Division	Physiographic Division	Rain Gauge ID
CPZ01	CD1 (Tidewater)	PGD1 (Coastal Plain)	446475, 448800, 449151
CPZ02	CD1 (Tidewater)	PGD2 (Piedmont)	448129
CPZ03	CD2 (Eastern Piedmont)	PGD1 (Coastal Plain)	None
CPZ04	CD2 (Eastern Piedmont)	PGD2 (Piedmont)	440778, 440993, 441322, 441929, 442941, 443192, 443200, 444414
CPZ05	CD3 (Western Piedmont)	PGD2 (Piedmont)	440166, 441614, 446178, 446692, 447025, 447338, 449272
CPZ06	CD3 (Western Piedmont)	PGD3 (Blue Ridge)	440561, 445690
CPZ07	CD4 (Northern)	PGD2 (Piedmont)	442159, 442729, 446712, 447130, 447164, 448396
CPZ08	CD4 (Northern)	PGD3 (Blue Ridge)	440720, 445851
CPZ09	CD4 (Northern)	PGD4 (Valley and Ridge)	442663, 443229, 448046, 448149
CPZ10	CD5 (Central Mountain)	PGD4 (Valley and Ridge)	442044, 442208, 443310, 444128, 445142, 445423, 445595, 445880, 448062, 448172, 449159
CPZ11	CD6 (Southwestern Mountain)	PGD3 (Blue Ridge)	443272, 444246, 448547, 449169
CPZ12	CD6 (Southwestern Mountain)	PGD4 (Valley and Ridge)	440766, 446955, 448022 449060, 449301
CPZ13	CD6 (Southwestern Mountain)	PGD5 (Appalachian Plateaus)	442269, 444180, 444410, 449215

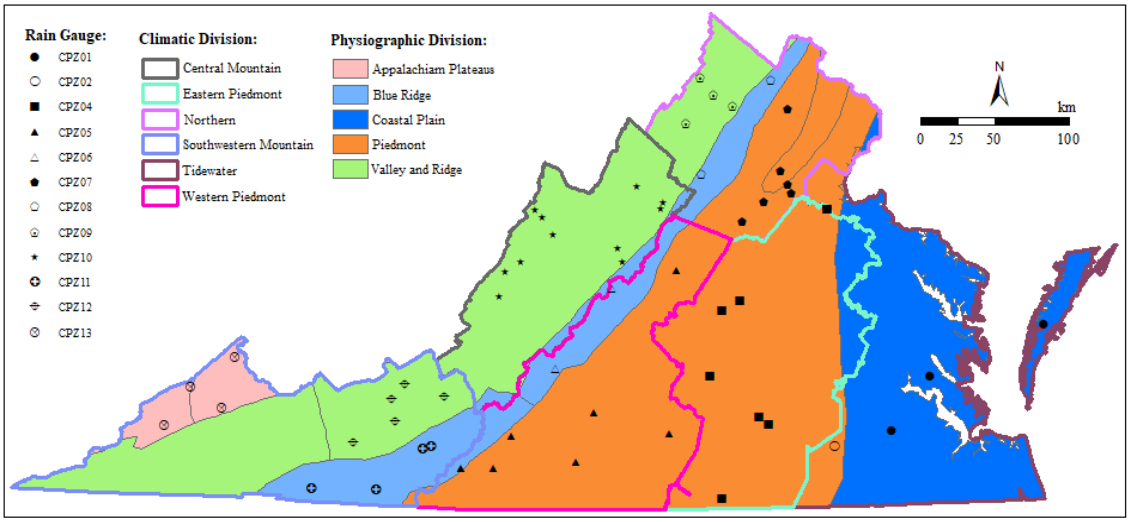

Figure 3.2. The Virginian climatic and physiographic divisions superimposed by 57 rain gauges with 15-min rainfall data.

3.2 Available Data

3.2.1 Rainfall Data

This thesis used 15-min rainfall data for the historical (prior 2013) periods of 57 rain gauges in Virginia and the projected precipitation time series by twelve pairs of Regional Climate Model (RCM) and Global Circulation Model (GCM). The data on 15-min precipitation observed at 57 rain gauges were downloaded from the NOAA National Climatic Data Center (NCDC) website (https://www.ncdc.noaa.gov/data-access/land-based-station-data). The gauges were grouped by the climatic-physiographic zones (Table 3.1 and Figure 3.2). Herein, it was hypothesized that the data at the gauges within a same zone are from a same population and can be pooled together into one dataset for statistical analysis. The rationale behind this hypothesis is that the spatial variability of precipitation across the zone might be statistically insignificant because each CPZ has a similar micro-climate and a similar physiology, as stated above. However, given the limited time, this project could not test this hypothesis using a statistical technique. A test will be done once a follow-up research will be awarded.

For a given rain gauge, the record only has times when precipitation was nonzero. To make the record consecutive at a 15-min interval, the times when precipitation was zero were added back by executing a Microsoft® Excel VBA program developed by the project team. In addition, the record has missing values for a time interval or more on a record day and/or for one day or more in a record year. The missing values were filled by executing another VBA program, which estimates a missing value as a function of the responding values at the neighboring gauges of this rain gauge. This function is expressed as:

$$P_{x,j} = \overline{P}_x \frac{\sum_{i=1}^{m} \frac{P_{i,j}}{\overline{P}_i}}{m} \quad \text{(Eq.3-1)}$$

where Px,j is the estimated 15-min precipitation of gauge x at time j; $\overline{P}_x$ is the mean annual precipitation of gauge x; Pi,j is the observed 15-min precipitation of gauge i at time j; $\overline{P}_i$ is the mean annual precipitation of gauge i; and m is the number of the neighboring gauges of gauge x. The mean annual precipitations of the 57 gauges, obtained from NOAA-NCDC and der Leeden (1994), are given in Table 3.2.

Table 3.2. The 57 rain gauges and their climatic-physiographic zones.

Gauge Name	ID	Begin Date	End Date	Elevation (m)	Mean Annual Precipitation (mm)[1]	Divisions	Zone
Painter 2 W	446475	05/02/1971	09/30/2012	9.1	1121.79	CD1⊗PGD1	CPZ01
CPZ01Wakefield 1 NW	448800	05/31/1985	02/28/2013	34.4	1204.72	CD1⊗PGD1	
Williamsburg 2 N	449151	05/02/1971	02/28/2013	21.3	1236.35	CD1⊗PGD1	
Stony Creek 2 N	448129	05/02/1974	04/30/1985	32.0	1193.04	CD1⊗PGD2	CPZ02
Blackstone Water Wor	440778	05/03/1971	03/31/1974	128.0	1133.95	CD2⊗PGD2	CPZ04
Bremo Bluff	440993	07/31/1986	02/28/2013	68.6	1087.75	CD2⊗PGD2	
Camp Pickett	441322	03/31/1974	02/28/2013	100.6	1169.42	CD2⊗PGD2	
Columbia 2 SSE	441929	05/07/1971	05/31/1986	88.4	1000.32	CD2⊗PGD2	
Farmville 2 N	442941	07/31/2009	12/31/2012	137.2	1126.24	CD2⊗PGD2	
Fredericksburg 2	443200	08/31/1978	02/28/1993	36.6	1044.51	CD2⊗PGD2	
Fredericksburg National Park	443192	05/02/1971	08/31/1978	27.4	1044.51	CD2⊗PGD2	
John H Kerr Dam	444414	05/07/1971	02/28/2013	76.2	1103.50	CD2⊗PGD2	
Altavista	440166	12/31/1983	02/28/2013	161.2	1118.11	CD3⊗PGD2	CPZ05
Chatham	441614	05/06/1971	02/28/2013	198.4	1149.60	CD3⊗PGD2	
North Garden	446178	05/31/1971	02/29/1992	209.1	1129.46	CD3⊗PGD2	
Philpott Dam 2	446692	05/06/1971	05/31/2009	342.3	1278.38	CD3⊗PGD2	

Randolph 5 NNE	447025	05/12/1971	01/31/1984	107.0	1077.36	CD3⊗PGD2	
Rocky Mount	447338	05/06/1971	02/28/2013	400.8	1189.74	CD3⊗PGD2	
Woolwine 4 S	449272	12/31/1983	02/28/2013	457.2	1293.88	CD3⊗PGD2	
Bedford	440561	01/31/1996	02/28/2013	374.3	1122.68	CD3⊗PGD3	CPZ06
Montebello Fish Hatchery	445690	04/30/1971	08/31/2007	812.9	1125.29	CD3⊗PGD3	
Culpeper Riverside Coast Guard	442159	07/01/1979	12/31/2003	79.2	1046.62	CD4⊗PGD2	CPZ07
Elkwood 6 SE	442729	06/04/1972	06/02/1984	100.0	1045.29	CD4⊗PGD2	
Piemont Research Station	446712	05/07/1971	02/28/2013	158.5	1112.90	CD4⊗PGD2	
Remington 2	447130	07/07/1979	02/28/1989	85.3	1120.14	CD4⊗PGD2	
Richardsville	447164	06/30/1984	04/30/1987	105.2	1044.08	CD4⊗PGD2	
The Plains 2 NNE	448396	05/01/1971	09/30/2004	161.5	1118.87	CD4⊗PGD2	
Big Meadows	440720	05/02/1971	07/31/1976	1079.0	1385.19	CD4⊗PGD3	CPZ08
Mount Weather	445851	05/02/1971	01/31/1987	505.7	1099.57	CD4⊗PGD3	
Edinburg	442663	06/30/1996	03/31/1999	282.9	896.11	CD4⊗PGD4	CPZ09
Front Royal	443229	01/01/1979	03/31/1990	283.5	1039.37	CD4⊗PGD4	
Star Tannery	448046	05/02/1972	01/31/2012	289.6	1023.62	CD4⊗PGD4	
Strasburg 2 ESE	448149	12/31/1978	04/30/1984	195.1	1068.64	CD4⊗PGD4	
Cobington Filter Plant	442044	12/31/1972	08/31/2011	374.9	952.75	CD5⊗PGD4	CPZ10
Dale Enterprise	442208	09/11/1978	01/31/2009	413.9	922.02	CD5⊗PGD4	
Gathright Dam	443310	12/31/1983	02/28/2013	539.5	986.54	CD5⊗PGD4	
Hot Springs	444128	09/04/1970	08/31/2011	681.5	1097.41	CD5⊗PGD4	
Lynnwood	445142	09/30/1983	12/01/1983	309.1	938.17	CD5⊗PGD4	
Mc Gaheysville 2 S	445423	04/30/1971	11/30/1983	331.9	1149.60	CD5⊗PGD4	
Millgap 2 NNW	445595	09/01/1976	02/28/2013	737.9	1124.71	CD5⊗PGD4	
Mustoe 1 SW	445880	06/30/1982	10/28/2007	725.4	1135.89	CD5⊗PGD4	
Staunton Water Treatment Plant	448062	12/31/1972	08/31/2007	51.5	989.96	CD5⊗PGD4	
Stuarts Draft	448172	05/01/1979	05/29/1984	442.0	1058.43	CD5⊗PGD4	
Williamsburg 2 S	449159	07/01/1978	08/31/2011	499.9	1029.32	CD5⊗PGD4	
Galax Water Plant	443272	04/01/1972	02/28/2013	719.3	1005.84	CD6⊗PGD3	CPZ11
Indian Valley	444246	04/30/1973	09/30/1993	823.0	1063.93	CD6⊗PGD3	
Trout Dale 3 SSE	448547	03/31/1974	02/28/2013	865.3	1077.98	CD6⊗PGD3	
Willis	449169	09/30/1993	02/28/2013	856.5	1144.78	CD6⊗PGD3	
Blacksburg National Weather	440766	03/31/2003	02/28/2013	604.1	1060.70	CD6⊗PGD4	CPZ12
Pulaski	446955	04/01/1972	02/28/2013	563.9	949.20	CD6⊗PGD4	
Staffordsville 3 ENE	448022	11/30/1993	02/28/2013	594.4	1000.25	CD6⊗PGD4	
White Gate	449060	12/31/1983	09/30/1993	563.9	965.88	CD6⊗PGD4	
Wytheville 1 S	449301	12/31/1983	02/28/2013	637.9	968.63	CD6⊗PGD4	
Davenport 2 NE	442269	12/31/1983	05/31/1986	488.0	1157.38	CD6⊗PGD5	CPZ13
Hurley 4 S	444180	03/31/1973	12/31/2009	331.6	1135.39	CD6⊗PGD5	
John Flannagan Lake	444410	12/31/1983	11/30/1991	445.0	1144.02	CD6⊗PGD5	
Wise 3 E	449215	12/31/1983	02/28/2013	776.9	1206.25	CD6⊗PGD5	

[1] The black number is the average of the values from https://www.ncei.noaa.gov/data/climate-normals-deprecated/access/clim20/va and https://www.ncdc.noaa.gov/cdo-web/datatools/normals and when both present the mean annual precipitations, whereas, it is the value from one of these two websites whichever presents the mean annual precipitation. On the other hand, the red number is from der Leeden (1994).

For a rain gauge, its neighboring gauges (Table 3.3) were selected as those that are within a geographic distance of 50 km and have relatively fewer missing values. In the table, a rain gauge of interest is highlighted in red, while its neighboring gauges are highlighted in black. At a time

when the precipitation at the rain gauge was filled, the neighboring gauges with observations were used in Eq. 1 and those without observations were excluded.

For each of the 57 rain gauges, the missing-filled 15-min precipitation time series was used to generate a dataset of annual maximum 15-min precipitation (designated $\tilde{X}_{15m,i}$ for description purpose, where subscript "15m" signifies the duration of 15 min; and i = 1, 2, …, 57, signifies the gauge). For a given observation year, the element value of $\tilde{X}_{15m,i}$ was computed as the maximum of the observed values (at gauge i) within this year. In addition, for each of the other eleven durations of longer than 15 min, the durational precipitation time series was formulated based on the missing-filled 15-min precipitation time series: the interval values of the durational time series was computed as accumulation (from beginning of the record) of the observed values of 15-min precipitation the duration apart. For instance, for the duration of 30 min, the first value of the 30-min precipitation time series was computed as the summation of first two observed values of the corresponding 15-min precipitation time series, the second value of the 30-min precipitation time series was computed as the summation of third and fourth observed values of the 15-min precipitation time series, and so on. For the duration of 72 h, the first value of the 72-h precipitation time series was computed as the summation of first 288 ($= 72 \times 60 \div 15$) values of the15-min precipitation time series. As a result, eleven more time series, which respectively have durations of 30 and 45 min and 1, 2, 3, 4, 6, 12, 24, 48 and 72 h, were formulated for the rain gauge. Further, for each of the eleven-time series and for a given observation year, the annual maximum durational precipitation is computed. This generated another eleven datasets of annual maximum durational precipitation for gauge i, namely $\tilde{X}_{30m,i}$, $\tilde{X}_{45m,i}$, $\tilde{X}_{1h,i}$, $\tilde{X}_{2h,i}$, $\tilde{X}_{3h,i}$, $\tilde{X}_{4h,i}$, $\tilde{X}_{6h,i}$, $\tilde{X}_{12h,i}$, $\tilde{X}_{24h,i}$, $\tilde{X}_{48h,i}$, and $\tilde{X}_{72h,i}$.

Table 3.3. Groups of the neighboring rain gauges (signified by their IDs) for filling missing values.[1]

Group 01: 443229, 448149, 448046, 442663, 445851, 442208, 440720, 446712, 442159, 447164, 442729, 447130, 445423	Group 02: 445851, 448149, 448046, 443229	Group 03: 442663, 448149, 448046, 443229, 440720, 442208	Group 04: 445142, 445423, 442208, 440720, 448062	Group 05: 440720, 445142, 445423, 446712
Group 06: 448062, 448172, 445690, 445142, 445423, 446178	Group 07: 446178, 445142, 448172, 448062, 440993, 445423, 445690, 441929	Group 08: 445690, 446178, 448172, 448062, 449159	Group 09: 445595, 445880, 449159, 443310, 444128,	Group 10: 442044, 443310, 444128, 445595, 445880, 445690, 449159
Group 11: 440766, 446955, 448022, 449060, 444246, 449169, 449301, 443272	Group 12: 449301, 449060, 446955, 443272, 448547	Group 13: 444246, 449169, 440766, 446955, 449272	Group 14: 443272, 446955, 444246, 449169, 449301, 448547	Group 15: 446692, 449272, 444246, 449169, 447338
Group 16: 447338, 446692, 449272, 441614	Group 17: 441614, 440166, 447338, 446692	Group 18: 440561, 440166, 447338	Group 19: 440166, 440561, 441614, 447338, 447025	Group 20: 448547, 446955, 449301, 443272
Group 21: 447025, 440778, 444414, 440166, 442941	Group 22: 440778, 441322, 447025, 444414, 442941, 448129	Group 23: 448129, 440778, 441322, 448800, 444414, 449151	Group 24: 448800, 440778, 441322, 448129, 449151, 444414	Group 25: 449151, 446475, 448129, 448800, 440778, 441322
Group 26: 446475, 449151	Group 27: 442159, 442729, 443192, 443200, 447130, 447164, 446712	Group 28: 446712, 442159, 440720, 442729, 447164, 447130, 440993	Group 29: 444414, 440778, 447025, 441322	Group 30: 442941, 447025, 440993, 441322
Group 31: 440993, 441929, 442941, 446178, 446712	Group 32: 442269, 444410, 444180, 449215, 446955, 449301, 448547	Group 33: 448396, 448149, 447130, 443229, 445851, 442729, 448046		

[1] In a given group, a rain gauge highlighted in red was filled by the other rain gauges of this group.

3.2.2 RCMs Predicted Precipitation

The predicted historic (i.e., pre-2013) and future (i.e., 2038 ~ 2070) data on regional precipitation at a 3-h time interval and a 50-km spatial resolution were downloaded from the North American Regional Climate Change Assessment Program (NARCCAP) website http://www.narccap.ucar.edu. To date, NARCCAP has generated twelve different dynamically downscaled datasets (Table 3.4), and all twelve datasets of precipitation for the grids (Figure 3.3) that cover Virginia were used in this study. The time series of precipitation were extracted from the NARCCAP ".nc" files using a computer program written in r language by the project team.

The extracted time series were stored in plain text files, which in turn were uploaded into Excel® spreadsheets for spatial and temporal downscaling.

Table 3.4. The twelve dynamically downscaled datasets by NARCCAP.

Dataset	Regional Climate Model (RCM)	General Circulation Model (GCM)
1	CRCM	CCSM
2		CGCM3
3	ECP2	GFDL
4		HadCM3
5	HRM3	GFDL
6		HadCM3
7	MM5I	CCSM
8		HadCM3
9	RCM3	CGCM3
10		GFDL
11	WRFG	CCSM
12		CGCM3

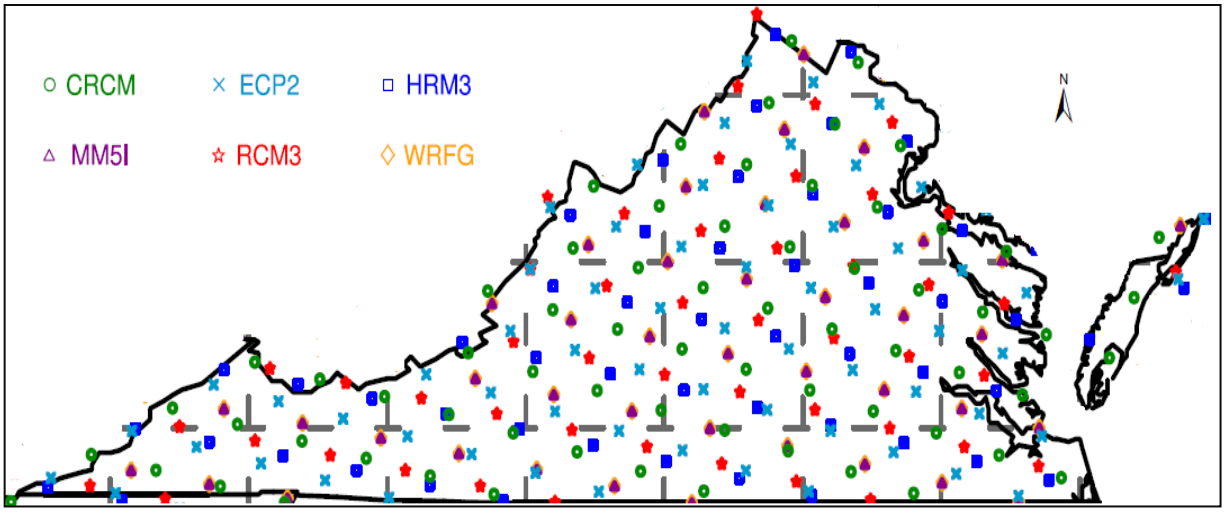

Figure 3.3. The centers of each 50-km grid cell for the six RCMs (not drawn to a scale). (Source: http://www.narccap.ucar.edu/data/gridpoint-maps.html).

3.3 Downscaling Methods

Downscaling of the RCMs' predictions was realized by four steps. First, the predicted 3-h precipitation at a rain gauge by an RCM (hereinafter referred to as RCM-predicted gauge precipitation) was computed as the inverse-distance-weighted average of the RCM's predictions for the four modeling grids surrounding the gauge. Second, the RCM-predicted gauge precipitation was corrected for possible errors in accordance with the observations at this same

gauge, deriving the spatially downscaled gauge precipitation. The equations used in this step were derived by regressing the observed annual maximum 3-h precipitations over the RCM-predicted annual maximum gauge 3-h precipitations for the historic period. Third, the spatially downscaled gauge precipitation was disaggregated to derive the 15-, 30- and 45-min and 1- and 2-h gauge precipitations. Fourth, the spatially downscaled gauge precipitation for a longer duration of 6, 12, 24, 48, or 72 h was derived from that for the duration of 3 h (i.e., the result of the second step). Herein, to alleviate the burden of data processing, the annual maximum durational precipitations rather than the time series were downscaled. The basic assumption is that the extreme values are independent of the mathematical operations involved in the downscaling procedure.

3.3.1 Spatial Downscaling

As shown in Figure 3.3, the centers of the 50-km grids are different for the six RCMs. For each RCM, its grid layer was overlaid with the layer of the rain gauges (Figure 3.2) in ArcMap® to identify the four grids surrounding each gauge. Subsequently, the geographic distances from the four grid centers to the rain gauge were calculated. At a given time, the RCM's predictions for the four grids were averaged using the inverses of the distances as the weights, resulting in a spatially averaged 3-h prediction. In addition, for each of the 57 rain gauges, the observed (formulated) 3-h time series $\tilde{X}_{3h,i}$ was regressed on the synchronic spatially averaged 3-h time series. That is, the regression was done for the record period of the gauge using the synchronically paired values of observed and spatially averaged 3-h precipitation. Further, taking the spatially averaged 3-h precipitations from 2038 to 2070, this regression equation was used to generate a 3-h precipitation time series at this gauge for this future period (i.e., downscale the projected precipitation). The regressions and computations were executed in Excel® 2010.

3.3.2 Temporal Downscaling

For each of the 57 rain gauges and for each of the six RCMs, the two spatially-downscaled 3-h precipitation time series, one for the record period and another for the future period, were used to generate two datasets of annual maximum 3-h precipitation, namely $\tilde{Y}_{3h,i}$ and $\tilde{Z}_{3h,i}$, respectively. To generate the annual maximum precipitation time series of 15, 30 and 45 min, and 1, 2 and 4 h (designated $\tilde{Y}_{15m,i}$, $\tilde{Y}_{30m,i}$, $\tilde{Y}_{45m,i}$, $\tilde{Y}_{1h,i}$, $\tilde{Y}_{2h,i}$, and $\tilde{Y}_{4h,i}$ for the record period, and $\tilde{Z}_{15m,i}$, $\tilde{Z}_{30m,i}$, $\tilde{Z}_{45m,i}$, $\tilde{Z}_{1h,i}$, $\tilde{Z}_{2h,i}$, and $\tilde{Z}_{4h,i}$ for the future period), $\tilde{X}_{15m,i}$, $\tilde{X}_{30m,i}$, $\tilde{X}_{45m,i}$, $\tilde{X}_{1h,i}$, $\tilde{X}_{2h,i}$, and $\tilde{X}_{4h,i}$ were separately regressed over $\tilde{X}_{3h,i}$ and then use $\tilde{Y}_{3h,i}$ and $\tilde{Z}_{3h,i}$ in the responding regression equations to get the responding time series. Herein, it was assumed that the regression equations were held regardless of the climatic conditions (Menabde et al., 1999; Socolofsky et al., 2001; Chang and Hiong, 2013; Mirhosseini et al., 2013). The regressions were implemented by each of the climatic-physiographic zones shown in Figure 3.2. For example, to regress $\tilde{X}_{15m,i}$ over $\tilde{X}_{3h,i}$ for CPZ01 (Table 3.1), the time series of $\tilde{X}_{15m,i}$ at the three rain gauges within this zone were pooled together into one 15-min dataset, while the time series of $\tilde{X}_{3h,i}$ at these same three rain gauges were pooled together into one 3-h dataset. To capsulate the datasets, if one value in the 3-h dataset corresponds to two or more values in the 15-min dataset, the arithmetic average, median, 75th percentile, and maximum of the multiple values were calculated, resulting in five capsulated datasets: one for 3-h and four for 15-min. The four capsulated 15-min datasets were separately regressed over the capsulated 3-h dataset. The regression equation with a largest coefficient of determination (R2) was chosen as the relationship between $\tilde{X}_{15m,i}$ and $\tilde{X}_{3h,i}$, and adopted to

generate $\tilde{Y}_{15m,i}$ and $\tilde{Z}_{15m,i}$. Both linear and nonlinear equations as well as piecewise regressions were tried to best fit the data.

On the other hand, to generate the annual maximum precipitation time series of 6, 12, 24, 48, and 72 h (designated $\tilde{Y}_{6h,i}$, $\tilde{Y}_{12h,i}$, $\tilde{Y}_{24h,i}$, $\tilde{Y}_{48h,i}$, and $\tilde{Y}_{72h,i}$ for the record period, and $\tilde{Z}_{6h,i}$, $\tilde{Z}_{12h,i}$, $\tilde{Z}_{24h,i}$, $\tilde{Z}_{48h,i}$, and $\tilde{Z}_{72h,i}$ for the future period), the accumulation procedure discussed in section 2.2.1 was applied to $\tilde{Y}_{3h,i}$ and $\tilde{Z}_{3h,i}$, respectively.

As a result, for each rain gauge and each RCM, twelve time series ($\tilde{Y}_{15m,i}$, $\tilde{Y}_{30m,i}$, $\tilde{Y}_{45m,i}$, $\tilde{Y}_{1h,i}$, $\tilde{Y}_{2h,i}$, $\tilde{Y}_{3h,i}$, $\tilde{Y}_{4h,i}$, $\tilde{Y}_{6h,i}$, $\tilde{Y}_{12h,i}$, $\tilde{Y}_{24h,i}$, $\tilde{Y}_{48h,i}$, and $\tilde{Y}_{72h,i}$) were generated for the record period and another twelve time series ($\tilde{Z}_{15m,i}$, $\tilde{Z}_{30m,i}$, $\tilde{Z}_{45m,i}$, $\tilde{Z}_{1h,i}$, $\tilde{Z}_{2h,i}$, $\tilde{Z}_{3h,i}$, $\tilde{Z}_{4h,i}$, $\tilde{Z}_{6h,i}$, $\tilde{Z}_{12h,i}$, $\tilde{Z}_{24h,i}$, $\tilde{Z}_{48h,i}$, and $\tilde{Z}_{72h,i}$) were generated for the future period. In total, 16,416 (= 57 gauges * 12 RCMs * 24 time series per gauge per RCM) datasets were generated. Again, all regressions and computations were executed in Excel® 2010.

3.3.2 Reliability Analyzing

We use all the 15-min rainfall data for the historical periods of 55 rain gauges (exclude station 442663 and 442941 for insufficient data) in Virginia to generate a box plot to analyze the reliability of the origin data (see Fig 3.4).

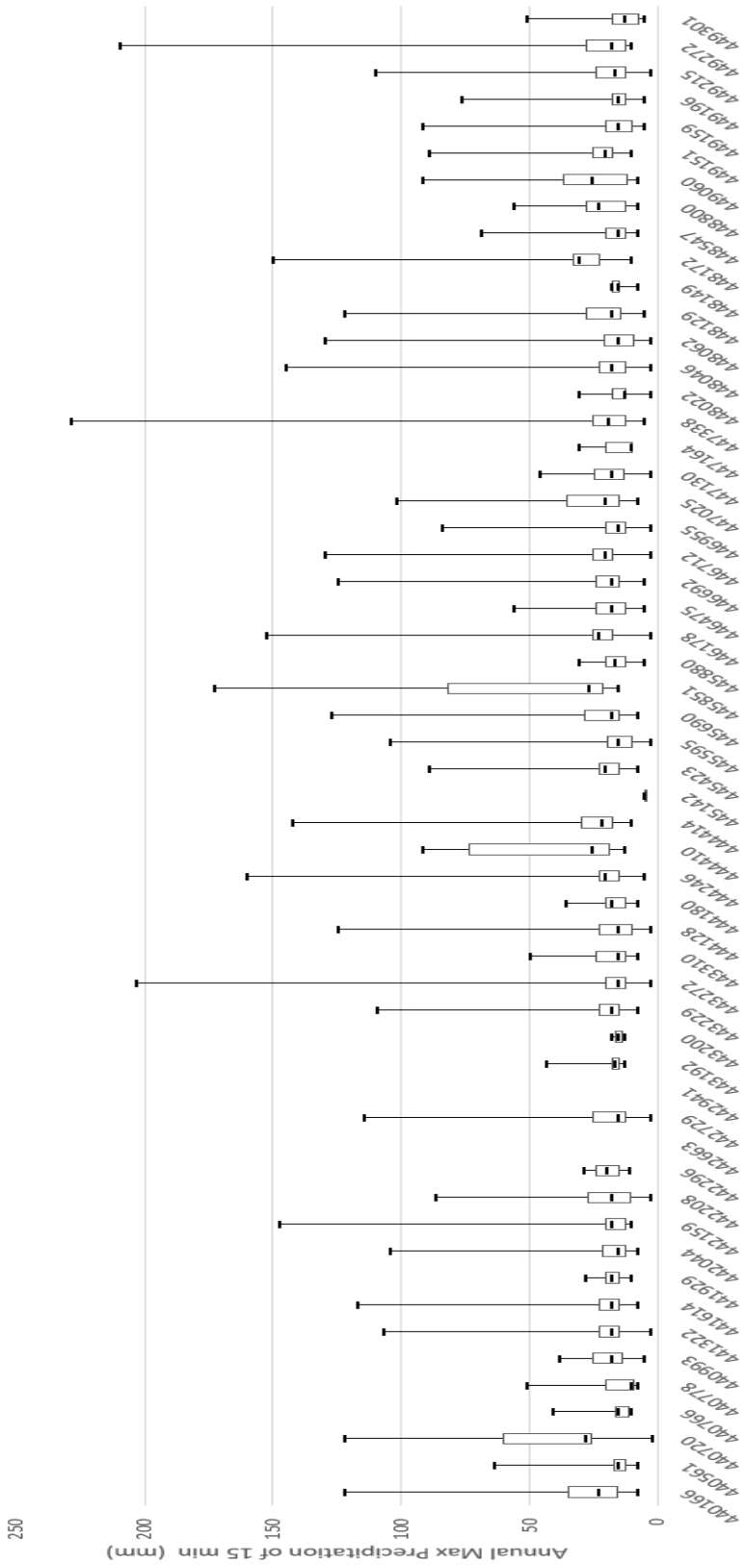

Fig 3.4 A box plot of max annual 15-min precipitation in 55 rain gauges

Among the 57 rain gauges, 25 of them were found to have a record period of 30 years or longer. According to Fig 3.4, the median values of annual maximum 15-min precipitation in all stations are under 50mm. Max values of annual maximum 15-min precipitation in station 443272, 444246, 445851, 446178 and 447338 are observed to be over 150mm, which means extraordinary rainfall events mainly happen in CPZ05 and CPZ08, the southern area of the Virginia along the Blue Ridge. On the other hand, strong rainfall events are relatively lacking in eastern Piedmont and Valley and Ridge (CPZ04, CPZ10, CPZ12 and CPZ13).

Generally, collected raw data shows no obvious contradiction with site topography, data is relatively reliable.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Downscaling Results

As discussed before, downscaling of the RCMs' predictions was realized by four steps. First, the predicted 3-h precipitation at a rain gauge by an RCM (hereinafter referred to as RCM-predicted gauge precipitation) was computed as the inverse-distance-weighted average of the RCM's predictions for the four modeling grids surrounding the gauge. Second, the RCM-predicted gauge precipitation was corrected for possible errors in accordance with the observations at this same gauge, deriving the spatially downscaled gauge precipitation. Third, the spatially downscaled gauge precipitation was disaggregated to derive the 15-, 30- and 45-min and 1- and 2-h gauge precipitations. Fourth, the spatially downscaled gauge precipitation for a longer duration of 6, 12, 24, 48, or 72 h was derived from that for the duration of 3 h (i.e., the result of the second step). Herein, to alleviate the burden of data processing, the annual maximum durational precipitations rather than the time series were downscaled. The basic assumption is that the extreme values are independent of the mathematical operations involved in the downscaling procedure.

The equations used in the second step (see Table 4.1) were derived by regressing the observed annual maximum 3-h precipitations over the RCM-predicted annual maximum gauge 3-h precipitations for the historic period, whereas, the equations used in the third step (see Table 4.2) were derived by regressing the observed annual maximum precipitations for a shorter or 4-h duration over those for the duration of 3 h. Both regressions were done by each of the climatic-physiographic zones. In this regard, the data at the rain gauges within a zone of interest were pooled together and then capsulated by excluding any abnormal and/or redundant values.

Table 4.1 The regression equations of RCM-predicted over observed 3-h annual maximum rainfall.

RCM	Zone	Regression Equation	Coefficient of Determination (R^2)
CRCM-CCSM	CPZ01	$P_{3h,obs} = 7.44(P_{3h,rcm}) - 44.76$	0.86
	CPZ02	$P_{3h,obs} = 6.82(P_{3h,rcm}) - 34.60$	0.87
	CPZ04	$P_{3h,obs} = 7.57(P_{3h,rcm}) - 45.09$	0.96
	CPZ05	$P_{3h,obs} = 10.01(P_{3h,rcm}) - 64.51$	0.87
	CPZ06	$P_{3h,obs} = 10.29(P_{3h,rcm}) - 61.20$	0.87
	CPZ07	$P_{3h,obs} = 11.59(P_{3h,rcm}) - 96.41$	0.93
	CPZ08	$P_{3h,obs} = 17.64(P_{3h,rcm}) - 153.41$	0.84
	CPZ09	$P_{3h,obs} = 3.74(P_{3h,rcm}) - 7.66 \quad P_{3h,rcm} \leq 15.0\ mm$	0.95
		$P_{3h,obs} = 64.29(P_{3h,rcm}) - 930.36 \quad P_{3h,rcm} > 15.0\ mm$	0.92
	CPZ10	$P_{3h,obs} = 7.18(P_{3h,rcm}) - 42.31$	0.88
	CPZ11	$P_{3h,obs} = 7.46(P_{3h,rcm}) - 37.43$	0.80
	CPZ12	$P_{3h,obs} = 6.47(P_{3h,rcm}) - 35.77$	0.90
	CPZ13	$P_{3h,obs} = 6.78(P_{3h,rcm}) - 42.87$	0.95

Table 4.2. The regression equations of observed shorter duration over 3-h annual maximum rainfall.

Zone	Duration	Regression Equation	Coefficient of Determination (R^2)
CPZ01	15 min	$P_{15m,obs} = 1.76(P_{3h,obs})^{0.62}$	0.67
	30 min	$P_{30m,obs} = 1.54(P_{3h,obs})^{0.73}$	0.83
	45 min	$P_{45m,obs} = 1.45(P_{3h,obs})^{0.78}$	0.89
	1 h	$P_{1h,obs} = 1.22(P_{3h,obs})^{0.86}$	0.95
	2 h	$P_{2h,obs} = 1.04(P_{3h,obs})^{0.96}$	0.99
	4 h	$P_{4h,obs} = 1.07(P_{3h,obs})^{1.01}$	0.99
CPZ02	15 min	$P_{15m,obs} = 0.25(P_{3h,obs})^{1.14}$	0.89
	30 min	$P_{30m,obs} = 0.21(P_{3h,obs})^{1.25}$	0.70
	45 min	$P_{45m,obs} = 0.55(P_{3h,obs})^{1.04}$	0.71
	1 h	$P_{1h,obs} = 0.36(P_{3h,obs})^{1.18}$	0.84
	2 h	$P_{2h,obs} = 0.46(P_{3h,obs})^{1.17}$	0.96
	4 h	$P_{4h,obs} = 1.01(P_{3h,obs})^{1.01}$	0.99
CPZ04	15 min	$P_{15m,obs} = 1.08(P_{3h,obs})^{0.76}$	0.90
	30 min	$P_{30m,obs} = 1.03(P_{3h,obs})^{0.86}$	0.97
	45 min	$P_{45m,obs} = 1.05(P_{3h,obs})^{0.89}$	0.98
	1 h	$P_{1h,obs} = 1.01(P_{3h,obs})^{0.92}$	0.99

	2 h	$P_{2h,obs} = 1.00\left(P_{3h,obs}\right)^{0.98}$	1.00
	4 h	$P_{4h,obs} = 1.16\left(P_{3h,obs}\right)^{0.99}$	0.99
CPZ05	15 min	$P_{15m,obs} = 1.14\left(P_{3h,obs}\right)^{0.75}$	0.62
	30 min	$P_{30m,obs} = 1.06\left(P_{3h,obs}\right)^{0.83}$	0.74
	45 min	$P_{45m,obs} = 0.98\left(P_{3h,obs}\right)^{0.89}$	0.85
	1 h	$P_{1h,obs} = 1.01\left(P_{3h,obs}\right)^{0.91}$	0.87
	2 h	$P_{2h,obs} = 0.83\left(P_{3h,obs}\right)^{1.02}$	0.99
	4 h	$P_{4h,obs} = 1.17\left(P_{3h,obs}\right)^{0.98}$	0.99
CPZ06	15 min	$P_{15m,obs} = 0.58\left(P_{3h,obs}\right)^{0.89}$	0.66
	30 min	$P_{30m,obs} = 0.84\left(P_{3h,obs}\right)^{0.88}$	0.74
	45 min	$P_{45m,obs} = 1.04\left(P_{3h,obs}\right)^{0.86}$	0.79
	1 h	$P_{1h,obs} = 0.99\left(P_{3h,obs}\right)^{0.90}$	0.84
	2 h	$P_{2h,obs} = 0.92\left(P_{3h,obs}\right)^{0.98}$	0.95
	4 h	$P_{4h,obs} = 1.25\left(P_{3h,obs}\right)^{0.97}$	0.98
CPZ07	15 min	$P_{15m,obs} = 0.56\left(P_{3h,obs}\right)^{0.93}$	0.64
	30 min	$P_{30m,obs} = 0.72\left(P_{3h,obs}\right)^{0.94}$	0.78
	45 min	$P_{45m,obs} = 0.68\left(P_{3h,obs}\right)^{0.99}$	0.88
	1 h	$P_{1h,obs} = 0.69\left(P_{3h,obs}\right)^{1.01}$	0.91
	2 h	$P_{2h,obs} = 0.40\left(P_{3h,obs}\right)^{1.18}$	0.94
	4 h	$P_{4h,obs} = 1.49\left(P_{3h,obs}\right)^{0.93}$	0.98
CPZ08	15 min	$P_{15m,obs} = 1.26\left(P_{3h,obs}\right)^{0.80}$	0.89
	30 min	$P_{30m,obs} = 1.13\left(P_{3h,obs}\right)^{0.86}$	0.95
	45 min	$P_{45m,obs} = 1.17\left(P_{3h,obs}\right)^{0.89}$	0.95
	1 h	$P_{1h,obs} = 1.17\left(P_{3h,obs}\right)^{0.90}$	0.96
	2 h	$P_{2h,obs} = 1.06\left(P_{3h,obs}\right)^{0.97}$	0.99
	4 h	$P_{2h,obs} = 1.18\left(P_{3h,obs}\right)^{0.99}$	0.95
CPZ09	15 min	$P_{15m,obs} = 0.78\left(P_{3h,obs}\right)^{0.85}$	0.55
	30 min	$P_{30m,obs} = 0.85\left(P_{3h,obs}\right)^{0.91}$	0.63
	45 min	$P_{45m,obs} = 0.73\left(P_{3h,obs}\right)^{0.98}$	0.72
	1 h	$P_{1h,obs} = 0.70\left(P_{3h,obs}\right)^{1.01}$	0.76
	2 h	$P_{2h,obs} = 0.76\left(P_{3h,obs}\right)^{1.03}$	0.94
	4 h	$P_{4h,obs} = 1.10\left(P_{3h,obs}\right)^{1.01}$	0.94
CPZ10	15 min	$P_{15m,obs} = 1.14\left(P_{3h,obs}\right)^{0.77}$	0.67
	30 min	$P_{30m,obs} = 0.79\left(P_{3h,obs}\right)^{0.93}$	0.81
	45 min	$P_{45m,obs} = 0.81\left(P_{3h,obs}\right)^{0.95}$	0.88
	1 h	$P_{1h,obs} = 0.77\left(P_{3h,obs}\right)^{0.98}$	0.88
	2 h	$P_{2h,obs} = 0.63\left(P_{3h,obs}\right)^{1.09}$	0.96
	4 h	$P_{4h,obs} = 1.18\left(P_{3h,obs}\right)^{0.98}$	0.98

CPZ11	15 min	$P_{15m,obs} = 0.65\left(P_{3h,obs}\right)^{0.88}$	0.67
	30 min	$P_{30m,obs} = 1.08\left(P_{3h,obs}\right)^{0.83}$	0.80
	45 min	$P_{45m,obs} = 1.06\left(P_{3h,obs}\right)^{0.88}$	0.88
	1 h	$P_{1h,obs} = 1.06\left(P_{3h,obs}\right)^{0.90}$	0.91
	2 h	$P_{2h,obs} = 1.17\left(P_{3h,obs}\right)^{0.93}$	0.96
	4 h	$P_{4h,obs} = 1.26\left(P_{3h,obs}\right)^{0.95}$	0.96
CPZ12	15 min	$P_{15m,obs} = 0.56\left(P_{3h,obs}\right)^{0.92}$	0.61
	30 min	$P_{30m,obs} = 0.82\left(P_{3h,obs}\right)^{0.90}$	0.78
	45 min	$P_{45m,obs} = 0.67\left(P_{3h,obs}\right)^{0.99}$	0.84
	1 h	$P_{1h,obs} = 0.69\left(P_{3h,obs}\right)^{1.01}$	0.89
	2 h	$P_{2h,obs} = 0.69\left(P_{3h,obs}\right)^{1.07}$	0.96
	4 h	$P_{4h,obs} = 1.42\left(P_{3h,obs}\right)^{0.93}$	0.98
CPZ13	15 min	$P_{15m,obs} = 0.65\left(P_{3h,obs}\right)^{0.92}$	0.62
	30 min	$P_{30m,obs} = 0.69\left(P_{3h,obs}\right)^{0.98}$	0.82
	45 min	$P_{45m,obs} = 0.59\left(P_{3h,obs}\right)^{1.05}$	0.89
	1 h	$P_{1h,obs} = 0.54\left(P_{3h,obs}\right)^{1.10}$	0.91
	2 h	$P_{2h,obs} = 0.84\left(P_{3h,obs}\right)^{1.02}$	0.94
	4 h	$P_{4h,obs} = 1.51\left(P_{3h,obs}\right)^{0.91}$	0.98

The equations used in the fourth step (see Table 4.3) were derived by regressing the RCM-predicted annual maximum gauge precipitations for a longer duration (i.e., 6, 12, 24, 48, or 72 h) over the RCM-predicted annual maximum gauge precipitations for one or more shorter durations spanning the entire RCM modeling period.

Table 4.3. The regression equations of RCM-predicted longer- over shorter-duration annual maximum rainfall.

Zone	Duration	Regression Equations	Coefficient of Determination (R^2)
CPZ01	6 h	$P_{6h,obs} = 1.854 + 1.530(P_{3h,obs})$	0.84
	12 h	$P_{12h,obs} = 4.084 + 1.266(P_{6h,obs})$	0.78
	24 h	$P_{24h,obs} = 3.870 + 1.207(P_{12h,obs})$	0.68
	48 h	$P_{48h,obs} = -0.997 + 1.266(P_{24h,obs})$	0.77
	72 h	$P_{72h,obs} = 1.922 + 1.073(P_{48h,obs})$	0.92
CPZ02	6 h	$P_{6h,obs} = 1.535 + 1.534(P_{3h,obs})$	0.85
	12 h	$P_{12h,obs} = 6.953 + 1.144(P_{6h,obs})$	0.80
	24 h	$P_{24h,obs} = 7.177 + 1.099(P_{12h,obs})$	0.70
	48 h	$P_{48h,obs} = 4.560 + 1.116(P_{24h,obs})$	0.69
	72 h	$P_{72h,obs} = 2.907 + 1.051(P_{48h,obs})$	0.92
CPZ04	6 h	$P_{6h,obs} = 1.872 + 1.494(P_{3h,obs})$	0.73
	12 h	$P_{12h,obs} = 4.803 + 1.241(P_{6h,obs})$	0.78
	24 h	$P_{24h,obs} = 7.670 + 1.073(P_{12h,obs})$	0.68
	48 h	$P_{48h,obs} = 5.770 + 1.094(P_{24h,obs})$	0.68
	72 h	$P_{72h,obs} = 1.671 + 1.081(P_{48h,obs})$	0.89
CPZ05	6 h	$P_{6h,obs} = 0.740 + 1.602(P_{3h,obs})$	0.77
	12 h	$P_{12h,obs} = 2.557 + 1.362(P_{6h,obs})$	0.81
	24 h	$P_{24h,obs} = 8.552 + 1.067(P_{12h,obs})$	0.70
	48 h	$P_{48h,obs} = 8.935 + 1.023(P_{24h,obs})$	0.65
	72 h	$P_{72h,obs} = 5.039 + 1.018(P_{48h,obs})$	0.85
CPZ06	6 h	$P_{6h,obs} = -0.569 + 1.738(P_{3h,obs})$	0.82
	12 h	$P_{12h,obs} = 3.004 + 1.332(P_{6h,obs})$	0.82
	24 h	$P_{24h,obs} = 9.135 + 1.033(P_{12h,obs})$	0.65
	48 h	$P_{48h,obs} = 7.652 + 1.034(P_{24h,obs})$	0.62
	72 h	$P_{72h,obs} = 7.421 + 0.966(P_{48h,obs})$	0.80
CPZ07	6 h	$P_{6h,obs} = 0.867 + 1.554(P_{3h,obs})$	0.74
	12 h	$P_{12h,obs} = 3.896 + 1.295(P_{6h,obs})$	0.78
	24 h	$P_{24h,obs} = 6.300 + 1.140(P_{12h,obs})$	0.67
	48 h	$P_{48h,obs} = -3.961 + 1.340(P_{24h,obs})$	0.76
	72 h	$P_{72h,obs} = 0.945 + 1.115(P_{48h,obs})$	0.91
CPZ08	6 h	$P_{6h,obs} = 4.249 + 1.324(P_{3h,obs})$	0.74
	12 h	$P_{12h,obs} = 4.937 + 1.237(P_{6h,obs})$	0.70
	24 h	$P_{24h,obs} = 6.353 + 1.149(P_{12h,obs})$	0.62
	48 h	$P_{48h,obs} = -0.244 + 1.258(P_{24h,obs})$	0.67
	72 h	$P_{72h,obs} = 2.249 + 1.088(P_{48h,obs})$	0.88
CPZ09	6 h	$P_{6h,obs} = 5.306 + 1.226(P_{3h,obs})$	0.72
	12 h	$P_{12h,obs} = 3.697 + 1.302(P_{6h,obs})$	0.76

	24 h	$P_{24h,obs} = 7.544 + 1.109(P_{12h,obs})$	0.67
	48 h	$P_{48h,obs} = 2.952 + 1.182(P_{24h,obs})$	0.63
	72 h	$P_{72h,obs} = 3.862 + 1.062(P_{48h,obs})$	0.86
CPZ10	6 h	$P_{6h,obs} = 1.555 + 1.551(P_{3h,obs})$	0.83
	12 h	$P_{12h,obs} = 4.470 + 1.264(P_{6h,obs})$	0.80
	24 h	$P_{24h,obs} = 7.913 + 1.065(P_{12h,obs})$	0.68
	48 h	$P_{48h,obs} = 7.181 + 1.047(P_{24h,obs})$	0.61
	72 h	$P_{72h,obs} = 6.315 + 0.982(P_{48h,obs})$	0.82
CPZ11	6 h	$P_{6h,obs} = 0.574 + 1.650(P_{3h,obs})$	0.87
	12 h	$P_{12h,obs} = 1.922 + 1.388(P_{6h,obs})$	0.86
	24 h	$P_{24h,obs} = 2.939 + 1.257(P_{12h,obs})$	0.74
	48 h	$P_{48h,obs} = 3.291 + 1.177(P_{24h,obs})$	0.74
	72 h	$P_{72h,obs} = 1.818 + 1.091(P_{48h,obs})$	0.91
CPZ12	6 h	$P_{6h,obs} = 0.145 + 1.693(P_{3h,obs})$	0.87
	12 h	$P_{12h,obs} = 2.798 + 1.336(P_{6h,obs})$	0.86
	24 h	$P_{24h,obs} = 3.461 + 1.213(P_{12h,obs})$	0.79
	48 h	$P_{48h,obs} = 6.856 + 1.077(P_{24h,obs})$	0.72
	72 h	$P_{72h,obs} = 8.380 + 0.961(P_{48h,obs})$	0.85
CPZ13	6 h	$P_{6h,obs} = 4.140 + 1.381(P_{3h,obs})$	0.70
	12 h	$P_{12h,obs} = 6.500 + 1.200(P_{6h,obs})$	0.63
	24 h	$P_{24h,obs} = 2.287 + 1.263(P_{12h,obs})$	0.75
	48 h	$P_{48h,obs} = 2.502 + 1.145(P_{24h,obs})$	0.74
	72 h	$P_{72h,obs} = 3.048 + 1.062(P_{48h,obs})$	0.83

The National Hydrography Dataset (NHD) (i.e., hydrography) was downloaded from the USGS website http://viewer.nationalmap.gov/viewer. NHD is a comprehensive set of digital spatial data about surface water features (e.g., rivers). This study used the 8-digit hydrologic cataloging units (HUCs or watersheds) presented by the NHD. Totally, there are 53 such HUCs in Virginia. In this chapter, we chose HUC2080202 (including 8 different stations from CPZ06 and CPZ10) as a sample to analyze our downscaling equations and results.

4.1.1 Spatial Downscaling

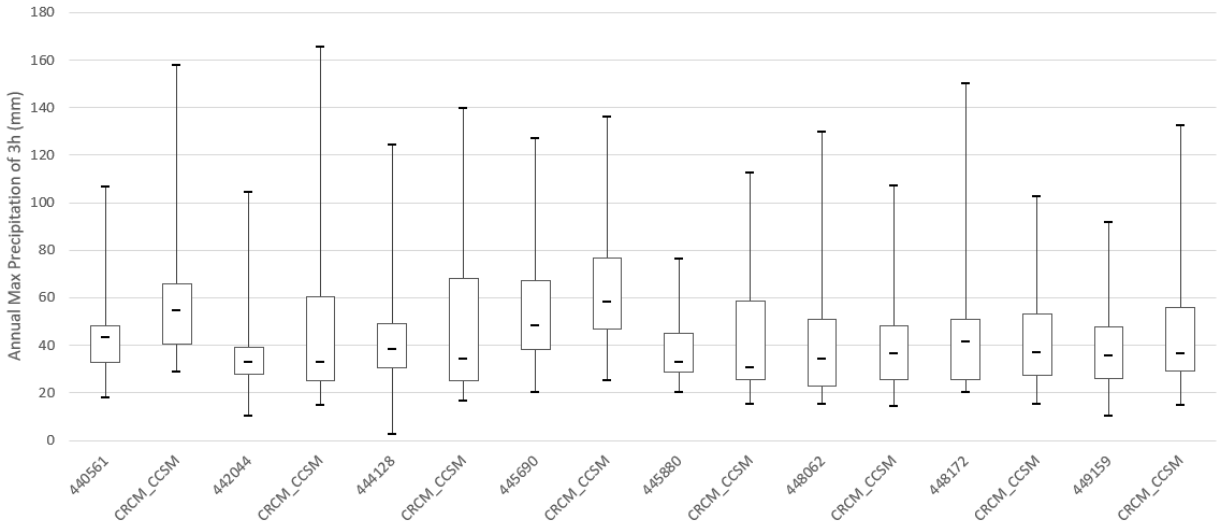

Fig 4.1 Spatial downscaling results in 8 different stations with 3h precipitation data.

Fig 4.1 shows the result of spatial downscaling in 8 different stations under CRCM_CCSM model with 3h precipitation data. According to the figure, the values from CRCM_CCSM have very similar median value with the observation data but cannot represent those extreme values (maximum and minimum values) that well. In most of the occasions, CRCM_CCSM model overestimates both maximum and minimum values and make the precipitation data more dispersed than history records but generally, this model successfully simulates the rainfall events in Virginia and can be used in the prediction of precipitation in the future for further studies of IDF curves.

4.1.2 Temporal Downscaling

Fig 4.2 shows the result of temporal downscaling in 8 different stations under CRCM_CCSM model with 15min precipitation data. According to Fig 4.1 and Fig 4.2, under a 15-min duration, the CRCM_CCSM model performed even better than under 3-h duration. The two datasets have not only very similar median value, but also have similar distribution. However, instead of being overestimated under 3-h duration, those extreme values (maximum and minimum values) have

been underestimated as an inadequacy under 15-min duration.

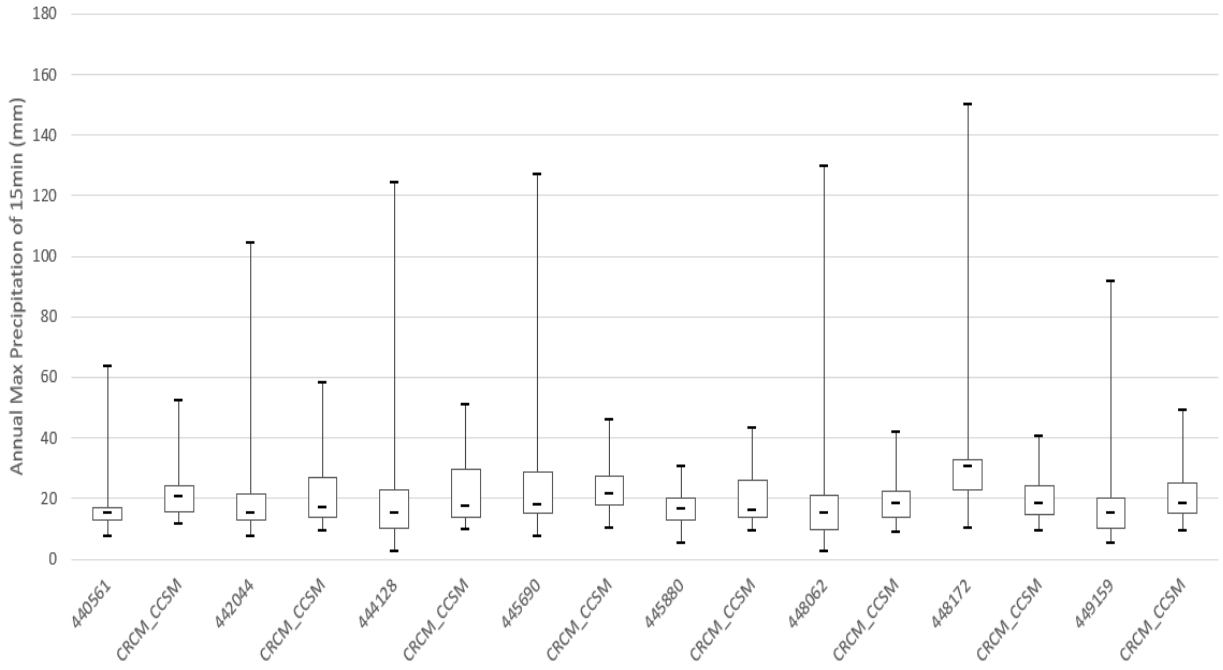

Fig 4.2 Spatial downscaling results in 8 different stations with 15min precipitation data.

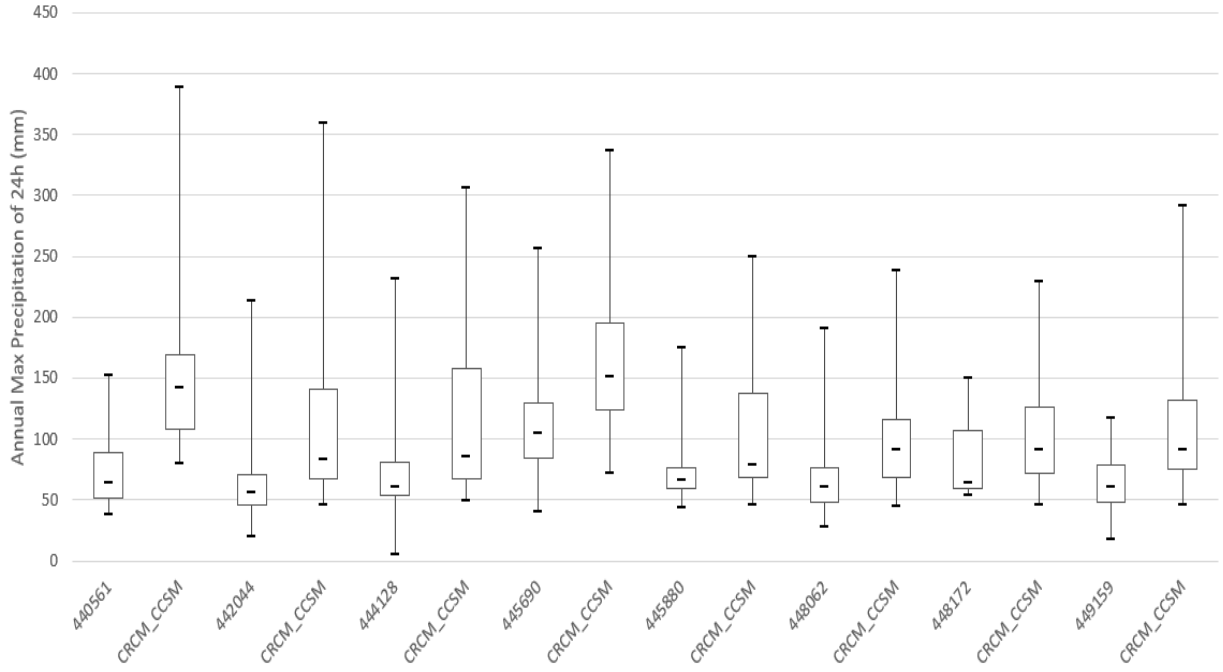

Fig 4.3 Spatial downscaling results in 8 different stations with 24h precipitation data.

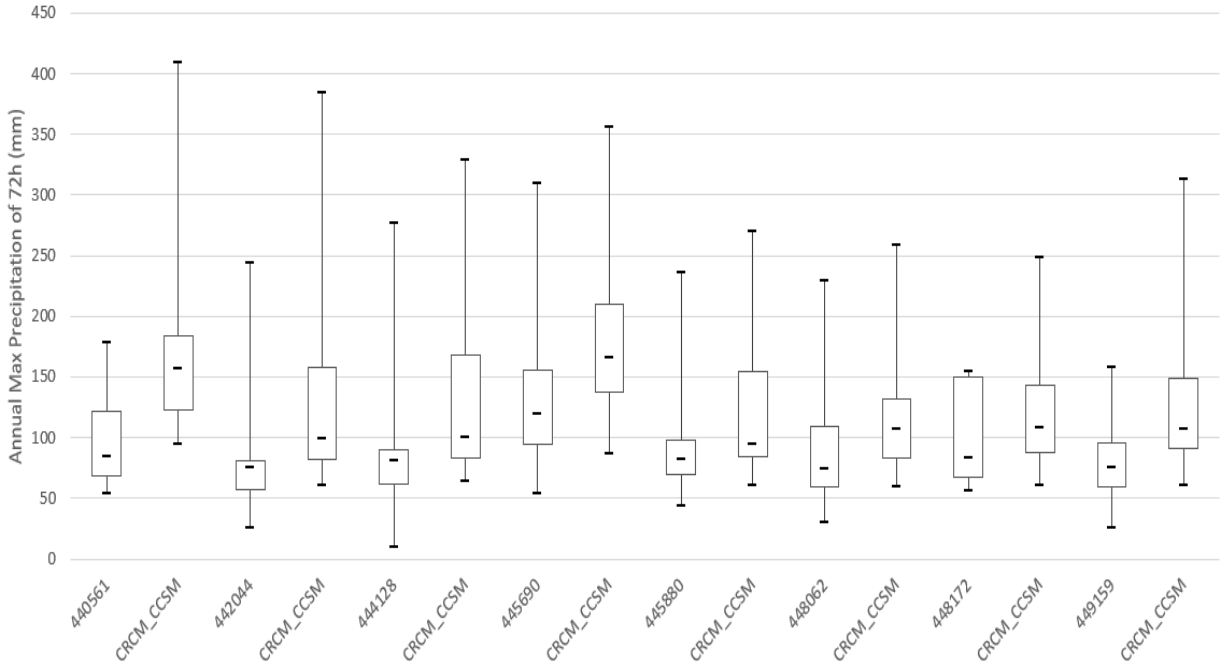

Fig 4.4 Spatial downscaling results in 8 different stations with 72h precipitation data.

Fig 4.3 and Fig 4.4 show the results of temporal downscaling in HUC2080202 under CRCM_CCSM model precipitation data with a duration of 24h and 72h. Comparing with 3-hour values, overestimations happened in not only representing those extreme values (maximum and minimum), but also in those ordinary events. All the precipitation values from the model have been deconcentrated and raised in varying degrees. In some stations (especially in station 440561) this phenomenon is particularly serious.

4.2 Results from Different RCMs

To compare the prediction results from different RCMs, we applied the equations from table 4.1 to all other 11 RCM-GCMs with the assumption that the equations were independent of the RCM-GCM models, trying to find whether they have statistical significance. We applied a significance level of $\alpha = 0.05/12=0.004167$ and did t-test between every two different RCMs. Table 4.4 shows the results. According to table 4.4 we can find that most RCMs have no

statistical significance under same RCM but different GCMs (only except CRCM). CRCM is a good RCM which is statistically significant from all other RCMs while RCM3_CGCM3 is only significant from CRCM models. Besides, most RCM-GCMs have 3 to 5 similar prediction models, which means practically we do not have to apply all these 12 models, several selected ones should be enough for further studies.

Table 4.4. The t-test results between every two RCM-GCMs.[1]

P value <=0.004167	CRCM_CCSM	CRCM_CGCM3	ECP2_GFDL	ECP2_HADCM3	HRM3_GFDL	HRM3_HADCM3	MM5I_GFDL	MM5I_HADCM3	RCM3_CGCM3	RCM3_GFDL	WRFG_CCSM	WRFG_CGCM3
CRCM_CCSM		0.002043	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000001	0.000028	0.000000	0.000000
CRCM_CGCM3	0.002043		0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000006	0.000543	0.000000	0.000000
ECP2_GFDL	0.000000	0.000000		0.641109	0.000021	0.000017	0.161884	0.000001	0.067605	0.357728	0.000017	0.000002
ECP2_HADCM3	0.000000	0.000000	0.641109		0.000088	0.000015	0.332813	0.000445	0.114740	0.193683	0.000200	0.000005
HRM3_GFDL	0.000000	0.000000	0.000021	0.000088		0.647501	0.000131	0.050733	0.021959	0.000011	0.096972	0.373576
HRM3_HADCM3	0.000000	0.000000	0.000017	0.000015	0.647501		0.001286	0.126684	0.029371	0.000029	0.226418	0.670690
MM5I_GFDL	0.000000	0.000000	0.161884	0.332813	0.000131	0.001286		0.004690	0.373884	0.052549	0.002313	0.000183
MM5I_HADCM3	0.000000	0.000000	0.000001	0.000445	0.050733	0.126684	0.004690		0.228415	0.000034	0.716130	0.174702
RCM3_CGCM3	0.000001	0.000006	0.067605	0.114740	0.021959	0.029371	0.373884	0.228415		0.029733	0.152764	0.047405
RCM3_GFDL	0.000028	0.000543	0.357728	0.193683	0.000011	0.000029	0.052549	0.000034	0.029733		0.000083	0.000010
WRFG_CCSM	0.000000	0.000000	0.000017	0.000200	0.096972	0.226418	0.002313	0.716130	0.152764	0.000083		0.288408
WRFG_CGCM3	0.000000	0.000000	0.000002	0.000005	0.373576	0.670690	0.000183	0.174702	0.047405	0.000010	0.288408	

[1] The black numbers mean the p values are smaller than α while red ones mean larger, in other words, have no statistical significance.

4.3 Discussion

The previous section reviewed the downscaling results of the CRCM_CCSM model. In the first two steps we corrected the possible statistical bias from the model and created our own version of the model that is well adjusted to a particular region. The main purpose of this step of spatial downscaling is to get a better match of the model output to the observations, which followed the suggestions from previous study (Brown et al., 2008). The results of the spatial downscaling generally responded as expected. According to the figures above, we can basically draw the conclusion that equations from table 4.1 can represent or predict precipitation eigenvalues with relative accuracy under a duration of 3 hours and performed well in adjusting to a different region.

The temporal downscaling results showed that CRCM_CCSM model can better reflect the historical values under shorter duration but overestimate the values under longer duration. This is under expectation for there are other studies that draw similar conclusions. Mirhosseini et al. (2013) used six different dynamically downscaled datasets to analyze the impact of climate change on IDF curves in Alabama. Their results revealed the precipitation pattern for short rainfall durations (i.e., less than 4 h), but for long durations (i.e., greater than 4 h) a large uncertainty on projected rainfall intensity made it difficult to draw any inclusive conclusions about expected changes of future rainfall intensity. That is, system bias or uncertainty accumulates over a long period of time and finally leads to significant difference between historical observations and model simulations.

Due to instrument accuracy and technical issues, a large amount of repeating data appeared in history records, observation data seemed to have a more concentrated distribution. However, predictions from RCMs cannot reflect this feature and after another process of temporal

downscaling using equations from table 4.2 and 4.3, system bias from downscaling could be cumulatively amplified and finally lead to the overestimations under longer durations. This phenomenon generally responds to an important consideration when interpreting fine-scale climate information mentioned in chapter 2: deriving climate projections at local scales is a multistep process and at each step, assumptions and approximations are made. Uncertainties are inherent in projections of changes in climate and their impacts. They arise from different sources and need to be kept in mind, whether explicitly quantified or not.

Furthermore, CRCM_CCSM cannot represent or predict those extreme rainfall events that precisely, which may have some influence on the further studies of IDF curves. In one step of generating IDF curves, the Fréchet, Weibull, and Gumbel distributions were tentatively fitted to the Weibull points for the whole record period as well as the periods of the sub-datasets. Based on the goodness of fit, we will select a best distribution for each rain gauge. Although individual extremum does not affect the distribution of the hole dataset, but it is possible that accidents happen. The resulting impacts should be kept in mind in follow-on steps.

Since uncertainty is inherent to the projections, an estimate of it should always be included and carried through the downscaling process. Such an estimate should at least include different potential future climate states and ideally should also estimate the influence of the downscaling procedure on the final results. That is the reason why in this study we took 12 different RCM-GCM models into account. In previous studies, Mirhosseini et al. (2013) used six different dynamically downscaled datasets (HRM3_HADCM3, CRCM_CGCM3, HRM3_GFDL, CRCM_CCSM, RCM3_GFDL, and ECP2_GFDL) to analyze the impact of climate change on IDF curves in Alabama but didn’t explain why excluded RCM3_CGCM3 and ECP2_HADCM3.

This can be explained by the t-test results that they have no statistical significance with RCM3_GFDL and ECP2_GFDL.

However, there is still another possibility that failures happen when equation sets for CRCM_CCSM are applied to other models. Usually, different RCMs contain distinct dynamical schemes and physical parameters, which means that RCMs driven by the same GCM can produce different results. However, we applied the equations from table 4.1 to all other 11 RCM-GCMs with the assumption that the equations were independent of the RCM-GCM models. Various RCMs differ in their numerical, physical, and technical aspects, it is reasonable that failures happen when equation sets for CRCM_CCSM are applied to other models. What's more, limited by time, table 4.2 showed results only from a single rain gauge, which can hardly represent the general situation. Thus, results from table 4.2 can only be considered as a very rough, beforehand criterion for computational efficiency concern. More related calculation is required in further steps of studies.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This thesis developed a statistical downscaling approach, which consists of a series of regression equations, to spatiotemporally downscale the RCMs' rainfall predictions in accordance with the observed 15-min rainfall data at the gauges in Virginia. From the development of this research thesis so far, we can basically draw the conclusion that the CRCM_CCSM model can better reflect the historical observations under a short duration but tends to overestimate the values under a long duration. The equations from Table 4.1 can be used to represent or predict precipitation eigenvalues with a good accuracy for a duration of 3 hours or shorter.

Because of the large amount of duplicate values in the history records, the observed data have a more concentrated distribution. However, the RCMs' predictions fail to reflect such a distribution and thus after another process of temporal downscaling using the equations from Tables 4.2 and 4.3, system biases might be cumulatively amplified, leading to the overestimations under longer durations. Furthermore, that the CRCM_CCSM model has a difficulty in precisely predicting some extreme rainfall events may somewhat influence practical applications of the model's predictions.

The CRCM is a distinctive RCM because its predictions are statistically different from the predictions of any other RCM models. In contrast, the predictions from the RCM3 model are statistically similar with those of other RCM models. Thus, to be cost-effective, practical studies may not need to examine the predictions of all twelve RCM models. It is likely good enough to use the predictions of seven of the models, namely the CRCM_CCSM, CRCM_CGCM, ECP2_HADCM3, HRM3_GFDL, MM5I_HADCM3, RCM3_GFDL and WRFG_CCSM.

5.2 Recommendations

The downscaling approach developed in this thesis has the capability of generating future precipitation data in the state of Virginia for watershed simulation and management and hydrologic engineering design. However, due to the time limitation, the downscaling equations were developed using the predictions of one of the twelve RCM models. Also, this thesis could not apply the developed equations to all rain gauges across the entire state. Future studies should verify and apply the equations for all climatic-physiographic zones of Virginia.

In terms of the existing results, while the simple liner regression shows its practicality for durations of 3 hours or shorter but tends to underestimate or overpredict some extreme precipitations for longer durations. This may be a drawback for some practical applications that need long-duration precipitations. In this regard, future studies should test nonlinear equations and/or more sophisticated algorithms.

With the assumption that the equations are independent of the RCM-GCM models, the twelve different RCM datasets generated by NARCCAP are found to have no necessity to take all of them into account in the subsequent computations. Chapter 4 applied the equations presented in Table 4.1 to the other eleven models. The technical concern of this assumption is that each of the models has different levels of simplification of the ocean-atmosphere-landscape physical processes as well as its own boundary conditions and limitations. Future studies should verify this assumption by redeveloping similar equations using these other models' predictions.

REFERENCES

Benestad, R. E. (2009). Downscaling precipitation extremes. *Theoretical and Applied Climatology,* 100(1-2), 1-21.

Bowden, Jared, Otte, Tanya, Nolte, Christopher, and Otte, Martin. (2012). Examining interior grid nudging techniques using two-way nesting in the WRF model for regional climate modeling." *Journal of Climate* 25.8: 2805-823.

Biau, Zorita, Von Storch, Wackernagel, and Biau, G. (1999). Estimation of precipitation by kriging in the EOF space of the sea level pressure field. *Journal of Climate* 12.4: 1070-085.

Brown, C., Greene, A. M., Block, P., & Giannini, A. (2008). Review of downscaling methodologies for Africa climate applications. IRI Technical Report, 08-05: IRI Downscaling Report, *International Research Institute for Climate and Society*, Columbia University.

Chang CW, and Hiong S. (2013). Estimation of sub-daily IDF curves in Singapore using simple scaling. In: Proceedings of International Conference on Climate Change Effects: Impacts World 2013. Potsdam, New York, NY, May 27-30.

Dickinson, R., Errico, R., Giorgi, F., & Bates, G. (1989). A regional climate model for the western United States. *Climatic Change*, 15(3), 383-422.

Emori, S., and S. J. Brown. (2005). Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophysical Research Letters*, 32.17 (2005).

Glahn, H., and D. Lowry. (1972). The use of model output statistics /MOS/ in objective weather forecasting. *Journal of Applied Meteorology* 11: 1203-211.

Hannachi, A. (2007). Pattern hunting in climate: a new method for finding trends in gridded climate data. *International Journal of Climatology* 27.1: 1-15.

Haylock, Malcolm R., Gavin C. Cawley, Colin Harpham, Rob L. Wilby, and Clare M. Goodess. (2006). Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology* 26.10: 1397-415.

Hertig, E., and J. Jacobeit. (2008). Downscaling future climate change: temperature scenarios for the Mediterranean area. *Global and Planetary Change* 63.2: 127-31.

Jeon, Chan-Hoo, Maarten C Buijsman, Alan J Wallcraft, Jay F Shriver, Brian K Arbic, James G Richman, and Patrick J Hogan. (2019). Improving surface tidal accuracy through two-way nesting in a global ocean model." *Ocean Modelling* 137: 98-113.

Johnson, Fiona, and Ashish Sharma. (2012). A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. Water Resources Research 48.1.

Kalnay, Eugenia, and Ming Cai. (2003). Impact of urbanization and land-use change on climate. *Nature* 528-31.

Kreft, Ita G. G, Jan De Leeuw, and Rien Van Der Leeden. (1994). Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. *The American Statistician* 48.4: 324-35

Lorenz, Philip, and D Jacob. Influence of regional scale information on the global circulation: a two‐way nesting climate simulation. *Geophysical Research Letters* 32.18 (2005).

Mailhot, A., S. Duchesne, D. Caya, and G. Talbot. (2007). Assessment of future change in intensity–duration–frequency (IDF) curves for southern Quebec using the Canadian Regional Climate Model (CRCM). *Journal of Hydrology*, 347 (1-2): 197-210.

Menabde M, Seed A, and Pegram G. (1999). A simple scaling model for extreme rainfall. *Water Resources Research* 35(1): 335-339.

Mirhosseini, G., P. Srivastava, and L. Stefanova. (2013). The impact of climate change on rainfall intensity–duration–frequency (IDF) curves in Alabama. *Regional Environmental Change*, 13(1): 25-33.

Nayak, Munir, and Ahmad Ghosh. (2013). Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. *Theoretical and Applied Climatology* 114.3-4: 583-603.

Palatella, Miglietta, Paradisi, and Lionello. (2010). Climate change assessment for Mediterranean agricultural areas by statistical downscaling. *Natural Hazards and Earth System Sciences* 10.7: 1647-1661.

Phillips, N., and J. Shukla. (1973). On the strategy of combining coarse and fine grid meshes in numerical weather prediction. *Journal of Applied Meteorology* 12: 763-70

Seaby, L. P., Refsgaard, J. C., Sonnenborg, T. O., Stisen, S., Christensen, J. H., & Jensen, K.H. (2013). Assessment of robustness and significance of climate change signals for an ensemble of distribution-based scaled climate projections. *Journal of Hydrology* 486(0), 479-493.

Semenov, M, and A. Barrow. (1997). Use of A stochastic weather generator in the development of climate change scenarios. *Climatic Change* 35.4: 397-414.

Socolofsky S, Adams E, and Entekhabi D. (2001). Disaggregation of daily rainfall for continuous watershed modeling. *Journal of Hydrologic Engineering* 6(4): 300-309.

Solomon, S, and Intergovernmental Panel on Climate Change. Climate Change 2007 the Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report

of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY, USA: IPCC.

The United States Agency for International Development, A review of downscaling methods for climate change projections (2014). Africa and Latin American Resilience to Climate Change Project. Retrieved from http://www.ciesin.org/documents/Downscaling_CLEARED_000.pdf

Tippett, M, T DelSole, S Mason, and A Barnston. (2008). Regression-based methods for finding coupled patterns. *Journal of Climate*: 4384-398.

Wilby, R.L., and T.M.L. Wigley. (2000). Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology* 20.6: 641-61.

Wilby, R. L., Troni, J., Biot, Y., Tedd, L., Hewitson, B. C., Smith, D. M., & Sutton, R. T. (2009). A review of climate risk information for adaptation and development planning. *International Journal of Climatology*, 29(9), 1193-1215.

Willems, P., and M. Vrac. (2011). Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *Journal of Hydrology* 402.3-4: 193-205.

Wu, R, S, Yang, S, Liu, L, Sun, Y, Lian, and Z, Gao. (2010). Changes in the relationship between northeast China summer temperature and ENSO. *Journal of Geophysical Research: Atmospheres* 115.D21.

Bio Sketch

Zhaoyi was born in 1995 in Fuzhou, China, with the family name of Cai. After completing his schoolwork at Fuzhou No.1 High School in 2013, Zhaoyi was enrolled into the Southeast University in Nanjing, China, in School of Transportation, majoring in City Underground Space Engineering. In 2017, he received a B.S. in Engineering. In August 2017, Zhaoyi began classes towards a M.S. degree in Civil Engineering at Old Dominion University (ODU).